DEX-0259

AA

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/11, 15/00, C07K 5/00, 16/00, C07H 21/04

A1

(11) International Publication Number: **WO 00/12694**

(43) International Publication Date: 9 March 2000 (09.03.00)

(21) International Application Number: PCT/US99/19508

(22) International Filing Date: 26 August 1999 (26.08.99)

(30) Priority Data:
60/098,183 26 August 1998 (26.08.98) US

(71) Applicant: MYRIAD GENETICS, INC. [US/US]; 320 Wakara Way, Salt Lake City, UT 84108 (US).

(72) Inventors: TAVTIGIAN, Sean; 557 East 1st Avenue, Salt Lake City, UT 84103 (US). TENG, David, H., F.; 1147 Blaine Avenue, Salt Lake City, UT 84105 (US). PERRY, William, L., III; 8940 Riverbend Ct., Indianapolis, IN 46250 (US). SCHROEDER, Marianne, M.; 3767 Redwood Circle, Palo Alto, CA 94306 (US). SIMARD, Jacques; 4808 DuPilet, St. Augustin de des Maures, Quebec G3A 129 (CA). ROMMENS, Johanna, M.; 1501 717 Bay Street, Toronto, Ontario M5G 2J9 (CA).

(74) Agents: IHNEN, Jeffrey, L. et al.; Rothwell, Figg, Ernst & Kurz, Suite 701 East, 555 13th Street N.W., Columbia Square, Washington, DC 20004 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CHROMOSOME 1-LINKED PROSTATE CANCER SUSCEPTIBILITY GENE AND MULTISITE TUMOR SUPPRESSOR

(57) Abstract

A human gene which is here named HPC1 has been identified in which mutations have been found which have been correlated with prostate cancer. The gene comprises 15 exons which have been sequenced. Transcripts of this gene have multiple splice variants which theoretically can lead to several thousand splice variants in accordance with certain rules which have been determined concerning the splicing of this gene's transcripts.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE OF THE INVENTION

CHROMOSOME 1-LINKED PROSTATE CANCER SUSCEPTIBILITY GENE AND MULTISITE TUMOR SUPPRESSOR

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC1), some mutant alleles of which cause susceptibility to cancer, in particular, prostate cancer. More specifically, the invention relates to germline mutations in the HPC1 gene and their use in the diagnosis of predisposition to prostate cancer. The present invention further relates to somatic mutations in the HPC1 gene in human prostate cancer and their use in the diagnosis and prognosis of human prostate cancer. Additionally, the invention relates to somatic mutations in the HPC1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC1 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving gain or loss of function of three loosely defined classes of genes: (1) dominant, positive regulators of the transformed state (oncogenes); (2) recessive, negative regulators of the transformed state (tumor suppressor genes); (3) recessive genes involved in maintenance of genome integrity (caretaker genes) (Kinzler and Vogelstein, 1997). Over one hundred oncogenes have been characterized. About a dozen tumor suppressor

and a similar number of caretaker genes have been identified; the number of genes falling into these last two classes is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of, human cancers. The most common oncogenic mutations are in the H-ras gene, found in 10-15% of all solid tumors (Anderson *et al.*, 1992). The most frequently mutated tumor predisposition genes are the TP53 gene, homozygously deleted or mutated in roughly 50% of all tumors, and CDKN2, which was homozygously deleted in 46% of tumor cell lines examined (Kamb *et al.*, 1994). Without a target that is common to all transformed cells, the dream of a "magic bullet" that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify oncogenes, tumor suppressor, and caretaker genes that play general roles in the process of oncogenesis.

Specific germline alleles of certain oncogenes, tumor suppressor, and caretaker genes are causally associated with predisposition to cancer. This set of genes is referred to as tumor predisposition genes. Some of the tumor predisposition genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms' tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); 9) Melanoma (CDKN2 and CDK4); 10) Breast and ovarian cancer (BRCA1 and BRCA2); 11) Cowden disease (MMAC1); 12) Multiple endocrine neoplasia (MEN1); 13) Nevoid basal cell carcinoma syndrome (PTC); 14) Tuberous sclerosis 2 (TSC2); 15) Xeroderma pigmentosum (genes involved in nucleotide excision repair); 16) Hereditary nonpolyposis colorectal cancer (genes involved in mismatch repair).

Tumor predisposition loci that have been mapped genetically but not yet isolated include genes for: Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); and Tuberous sclerosis 1 (TSC1). Tumor predisposition genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2

and CDK4), tyrosine phosphatases (MMAC1), as well as others with no obvious similarity to proteins of known function (BRCA2).

In many cases, the tumor predisposition gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration, whether germline, in tumors, or in tumor cell lines, may signify the position of important tumor predisposition genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor and caretaker genes characterized to date is that their function is lost at high frequency in certain tumor types. Loss of function is often a consequence of deletion. The deletions can involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation or because of a secondary sporadic mutation. Conversely, a number of oncogenes are subject to gain of function in certain tumor types. Gain of function often involves amplification of the copy number of a gene but may also occur when a chromosomal translocation generates a chimeric gene. Gain of function can also be a consequence of point mutations that alter some aspect of gene regulation or function.

Prostate cancer is the most common cancer in men in many western countries, and the second leading cause of cancer deaths in men. It accounts for more than 40,000 deaths in the US annually. The number of deaths is likely to continue rising over the next 10 to 15 years. In the US, prostate cancer is estimated to cost $1.5 billion per year in direct medical expenses. In addition to the burden of suffering, it is a major public-health issue. Numerous studies have provided evidence for familial clustering of prostate cancer, indicating that family history is a major risk factor for this disease (Cannon et al., 1982; Steinberg et al., 1990; Carter et al, 1993).

Prostate cancer has long been recognized to be, in part, a familial disease. Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci. Woolf (1960), described a relative risk of 3.0 of developing prostate cancer among first-degree relatives of prostate cancer cases in Utah using death certificate data. Relative risks ranging from 3 to 11 for first-degree relatives of prostate cancer cases have been reported (Cannon et al., 1982; Woolf, 1960; Fincham et al., 1990; Meikle et al., 1985; Krain, 1974; Morganti et al., 1956; Goldgar et al., 1994). Carter et al. (1992) performed segregation analysis on families ascertained

through a single prostate cancer proband. The analysis suggested Mendelian inheritance in a subset of families through autosomal dominant inheritance of a rare (q=0.003), high-risk allele with estimated cumulative risk of prostate cancer for carriers of 88% by age 85. Inherited prostate cancer susceptibility accounted for a significant proportion of early-onset disease, and overall was responsible for 9% of prostate occurrence by age 85. Recent results demonstrate that at least two loci exist which convey susceptibility to prostate cancer as well as other cancers. These loci are HPC1 on chromosome 1, (Smith et al., 1996), and one or more loci responsible for the unmapped residual.

Smith et al., (1996) indicated that the inherited prostate susceptibility in kindreds with early age onset is linked to chromosome 1. Most strategies for cloning the chromosome 1-linked prostate cancer predisposing gene (HPC1) require precise genetic localization studies. The simplest model for the functional role of HPC1 holds that alleles of HPC1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type HPC1 allele are not cancerous. However, cells that contain one wild type HPC1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of HPC1 and may develop into tumors. According to this model, predisposing alleles of HPC1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: men who possess one predisposing allele (and one wild type allele) risk developing cancer, because their prostate cells may spontaneously lose the wild type HPC1 allele. This model applies to both tumor suppressor and caretaker genes described above. By inference this model may also explain the HPC1 function, as has recently been suggested (Smith *et al.*, 1996).

A second possibility is that HPC1 predisposing alleles are truly dominant; that is, a wild type allele of HPC1 cannot overcome the tumor-forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of HPC1 before giving rise to malignant cells. Instead, prostate cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If HPC1 predisposing alleles are recessive, the HPC1 gene is expected to be expressed in normal prostate tissue but not functionally expressed in prostate tumors. In contrast, if HPC1 predisposing alleles are dominant, the wild type HPC1 gene may or may not be expressed in

normal prostate tissue. However, the predisposing allele will likely be expressed in prostate tumor cells.

Evidence for a prostate cancer susceptibility locus (HPC1) on the long arm of chromosome 1, which is hypothesized to explain approximately 35% of families, was recently presented (Smith et al., 1996). Although several groups report evidence supporting this localization, it has not yet been confirmed statistically. Both the original Smith et al. report and a subsequent analysis of additional families (Cooney et al., 1997), suggest that the bulk of linkage evidence comes from African- American high-risk kindreds. In addition, it appears that this gene predisposes (although not exclusively) primarily to early onset prostate cancer.

The chromosome 1 linkage of HPC1 has not been statistically confirmed; however, a report by Cooney et al. (1997) as well as a manuscript in review (Neuhausen et al., in press) are suggestive of confirmation, with less-than-significant indications of linkage at the location suggested to harbor HPC1. We have localized the HPC1 gene to an approximately 1 cM interval bounded by J23 and E15. A multipoint analysis of all 22 Utah kindreds resulted in a heterogeneity lod score of +1.20 at D1S254, with an estimate that 5% of kindreds are linked. This analysis excluded linkage for an alpha greater than .33. We have a set of 5 Utah kindreds showing evidence of a common segregating haplotype surrounding D1S254 which themselves define a region of less than 1cM in which HPC1 must lie.

Identification of a prostate cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC1), some alleles of which cause susceptibility to cancer, in particular prostate cancer. More specifically, the present invention relates to germline mutations in the HPC1 gene and their use in the diagnosis of predisposition to prostate cancer. The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPC1 gene. The invention further relates to somatic mutations in the HPC1 gene in human

prostate cancer and their use in the diagnosis and prognosis of human prostate cancer. Additionally, the invention relates to somatic mutations in the HPC1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC1 gene, (including gene therapy, protein replacement therapy. protein mimetics, and inhibitors). The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPS gene. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC1 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram showing the order of genetic markers neighboring HPC1, a schematic map of YACs spanning the HPC1 region, a schematic map of BACs and PACs spanning the HPC1 region, and also shows the location of the HPC1 gene within the genetically defined interval.

Figure 2 is a diagram of the HPC1 transcription unit showing the locations of the exons of HPC1 relative to the BAC/PAC contig and relative to each other. The individual exons are numbered, and these numbers correspond to their SEQ ID NOs.

Figure 3 shows multipoint Lod scores for the prostate cancer susceptibility locus relative to the 1q24-25 markers. A walking three point analysis with markers D1S2883, D1S254 and D1S412 is plotted as a function of distance from D1S2883. The combined values are plotted assuming all kindreds are linked, and with a heterogeneity estimate of 0.05 (estimated from HOMOG). The maximum Lod score under heterogeneity is 1.20 at D1S254.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC1), some alleles of which cause susceptibility to cancer, in particular prostate cancer. More specifically, the present invention relates to germline mutations in the HPC1 gene and their use in the diagnosis of predisposition to prostate cancer. The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPC1 gene. The invention further relates to somatic mutations in the HPC1 gene in human

prostate cancer and their use in the diagnosis and prognosis of human prostate cancer. Additionally, the invention relates to somatic mutations in the HPC1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC1 gene, (including gene therapy, protein replacement therapy. protein mimetics, and inhibitors). The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPS gene. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC1 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the HPC1 locus or of a mutated HPC1 locus, preferably at least eight bases and not more than about 300 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the HPC1 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the HPC1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the HPC1 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the HPC1 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the HPC1 locus, the kits comprising a polynucleotide complementary to the portion of the HPC1 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the HPC1 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the HPC1 locus.

The present invention further provides methods of screening the HPC1 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the HPC1

locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the HPC1 locus. Such methods may also include a step of providing the complete set of short polynucleotides defined by the sequence of HPC1 or discrete subsets of that sequence, all single-base substitutions of that sequence or discrete subsets of that sequence, all 1-, 2-, 3-, or 4-base deletions of that sequence or discrete subsets of that sequence, and all 1-, 2-, 3-, or 4-base insertions in that sequence or discrete subsets of that sequence. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected HPC1 mutant alleles to identify mutations in the HPC1 gene.

In addition, the present invention provides methods to screen drugs for inhibition or restoration of HPC1 gene product function as an anticancer therapy.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the HPC1 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the HPC1 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of HPC1. These may functionally replace the activity of HPC1 *in vivo*.

It is a discovery of the present invention that the HPC1 locus which predisposes individuals to prostate cancer, is a gene encoding an HPC1 protein, which has been found to have no significant homology with publicly available protein or DNA sequences. This gene is termed HPC1 herein. It is a discovery of the present invention that mutations in the HPC1 locus in the germline are indicative of a predisposition to prostate cancer. It is a discovery of the present invention that somatic mutations in the HPC1 locus are also associated with prostate and other types of cancer. Finally, it is a discovery of the present invention that two common polymorphisms of HPC1 are associated with both prostate and many other types of cancer. The mutational events of the HPC1 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

STRATEGY FOR THE MOLECULAR CLONING OF HPC1

Starting from a region on chromosome 1 of the human genome, a region which contains a genetic locus, HPC1, which causes susceptibility to cancer, including prostate cancer, has been identified.

The region containing the HPC1 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the HPC1 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of prostate cancer, a chromosomal region has been pinpointed that contains the HPC1 gene as well as putative susceptibility alleles in the HPC1 locus. Two meiotic breakpoints have been discovered on the distal side of the HPC1 locus which are expressed as recombinants between genetic markers and the disease, and one recombinant on the proximal side of the HPC1 locus. Thus, a region which contains the HPC1 locus is physically bounded by these markers.

Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently gives evidence whether or not an HPC1 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the HPC1 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying prostate cancer, since penetrance of the HPC1 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein *et al.*, 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys *et al.*, 1985; Nakamura *et al.*, 1987), and an abundant class of DNA polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt *et al.*, 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an *ad hoc* basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the target nucleic acid sequence using PCR; such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the HPC1 locus, then replacing these flanking markers with other markers that were successively closer to the HPC1 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the HPC1 locus as either distal or proximal to regionally localized specific genetic markers.

Contig assembly

Given a genetically defined interval flanked by meiotic recombinants, one needs to generate a contig of genomic clones that spans that interval. Publicly available resources, such as the Whitehead integrated maps of the human genome (e.g., the WICGR Chr 1 map of Nov. 19, 1996) provide aligned chromosome maps of genetic markers, other sequence tagged sites (STSs), radiation hybrid map data, and CEPH yeast artificial chromosome (YAC) clones. From the map data, one can often identify a set of yeast artificial chromosomes (YACs) that span the genetically defined interval. Oligonucleotide primer pairs for the markers located in the interval can be synthesized and used to screen libraries of bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs). Successive rounds of BAC/PAC library screening with BAC

or PAC end markers enables the completion of a BAC/PAC clone contig that spans the genetically defined interval. A set of overlapping but non-redundant BAC and PAC clones that span this interval (Figure 1C) (the tiling path) can then be selected for use in subsequent molecular cloning protocols.

Genomic sequencing

Given a tiling path of BAC and PAC clones across a defined interval, one useful gene finding strategy is to generate an almost complete genomic sequence of that interval. Random genomic clone sublibraries can be prepared from each BAC or PAC clone in the tiling path. Individual sublibrary clones sufficient in number to generate an, on average, 6x redundant sequence of each BAC or PAC can then be end-sequenced with vector primers. These sequences can be assembled into sequence contigs, and these contigs placed in a local genomic sequence database. One can search the genomic sequence contigs for sequence similarity with known genes and expressed sequence tags (ESTs), examine them for the presence of long open translational reading frames, and characterize them for CpG dinucleotide frequency.

Hybrid selection

Given a tiling path of BAC and PAC clones across a defined interval, another useful gene finding strategy is to obtain cDNA clones cognate to the tiling path BACs and PACs. One preferred cDNA cloning strategy is hybrid selection. cDNA can be prepared from a number of human tissues and human cell lines in such a manner that the cDNA molecules have PCR primer binding sites (anchors) at each end. This cDNA can be affinity captured with the tiling path BACs and PACs. Captured cDNA can then be PCR amplified using the anchor primers and then cloned. Individual clones can then be end-sequenced with vector primers. The sequences of these cDNA clones can be analyzed for similarity to genomic sequence contigs generated from BACs and PACs on the tiling path. One can then identify individual exons of genes in the genetically defined interval by parsing the sequences of true-positive hybrid selected clones across these genomic sequence contigs.

RACE and inter-exon PCR

While hybrid selection is an efficient approach to the initial identification of novel genes located within a defined interval of the genome, the approach is not often the most efficient way to complete the cloning of those genes. Rapid amplification of cDNA ends (RACE) provides a PCR based method to identify new 5' and 3' cDNA sequences. cDNA can be prepared from a number of human tissues in a manner such that the cDNA molecules have PCR primer binding

sites (anchors) at their 5' ends, 3' ends, or both. PCR amplification from this cDNA with 5' end anchor primers and gene specific reverse primers can generate 5' RACE products. Similarly, PCR amplification with 3' end anchor primers and gene specific forward primers can generate 3' RACE products. cDNA cloning techniques can also miss exons that lie between already known exons of a gene; for instance, this can easily occur if a particular exon is only included in a relatively rare splice variant of a transcript. Combinatorial inter-exon PCR is an effective strategy for detecting these exons. One can design a forward primer based on sequences from the first known exon of the gene and a set of reverse primers, one based on the sequence of each of the downstream exons (or any subset thereof) of the gene. Then one can PCR amplify from cDNA of tissues and cell lines thought to express the gene, using all the combinations of the forward primer with each reverse primer. Combinations as complex as a forward primer from each exon paired with a reverse primer from each exon, subject only to the limitation that the forward primer should be from an exon upstream of the exon from which the reverse primer was designed, can be tried. PCR products which differ in length from the expected product can be gel purified. In either RACE or combinatorial inter-exon PCRs, the PCR products can either be gel purified and then sequenced directly or first cloned and then sequenced.

cDNA library screening

Another useful strategy for finding new 5', 3', or internal sequences is cDNA library screening. One can make or purchase bacteriophage l cDNA libraries prepared from RNA from tissues or cell lines thought to express the gene. One then screens plaque lifts from those libraries with labeled nucleic acid probes based on the currently known sequences of the gene of interest. Individual positive clones are purified, and then the clone inserts can be sequenced.

Mutation screening

Proof that any particular gene located within the genetically defined interval is HPC1 is obtained by finding sequences in DNA or RNA extracted from affected kindred members which create abnormal HPC1 gene products or abnormal levels of HPC1 gene product. Such HPC1 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with prostate cancer than in individuals in the general population. Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline HPC1 alleles mutated into sequences which are identical or similar to HPC1 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing HPC1 sequences from tumor

tissue to HPC1 alleles from the germline of the same individuals, or one is comparing germline HPC1 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Small deletions or base pair substitutions could also significantly alter protein expression by changing the level of transcription, splice pattern, mRNA stability, or translation efficiency of the HPC1 transcript. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

Useful Diagnostic Techniques

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type HPC1 locus is detected. In addition, the method can be performed by detecting the wild-type HPC1 locus and confirming the lack of a predisposition to cancer at the HPC1 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of HPC1 mutations thus provides both diagnostic and prognostic information. An HPC1 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying an HPC1 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the HPC1 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the

promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to reduction or loss of expression of the HPC1 gene product, expression of an altered HPC1 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent *in situ* hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below. Also useful is the recently developed technique of DNA microchip technology.

Predisposition to cancers, such as prostate cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the HPC1 gene. For example, a person who has inherited a germline HPC1 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the HPC1 gene. Alteration of a wild-type HPC1 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as HPC1, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita *et al.*, 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield *et al.*, 1991), heteroduplex analysis (HA) (White *et al.*, 1992) and chemical mismatch cleavage (CMC) (Grompe *et al.*, 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory

mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type HPC1 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

Detection of point mutations may be accomplished by molecular cloning of the HPC1 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single-stranded conformation analysis (SSCA) (Orita *et al.*, 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell *et al.*, 1990; Sheffield *et al.*, 1989); 3) RNase protection assays (Finkelstein *et al.*, 1990; Kinszler *et al.*, 1991); 4) allele-specific oligonucleotides (ASOs) (Conner *et al.*, 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano and Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular HPC1 mutation. If the particular HPC1 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton *et al.*, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an

insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the HPC1 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is complementary to the human wild-type HPC1 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the HPC1 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the HPC1 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton *et al.*, 1988; Shenk *et al.*, 1975; Novack *et al.*, 1986.

Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the HPC1 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the HPC1 gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the HPC1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length (although shorter and longer oligomers are also usable as well recognized by those of skill in the art), corresponding to a portion of the HPC1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the HPC1 gene. Hybridization of allele-specific probes with amplified HPC1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under high stringency hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The newly developed technique of nucleic acid analysis via microchip technology is also applicable to the present invention. In this technique, literally thousands of distinct oligonucleotide probes are built up in an array on a silicon chip. Nucleic acid to be analyzed is fluorescently labeled and hybridized to the probes on the chip. It is also possible to study nucleic acid-protein interactions using these nucleic acid microchips. Using this technique one can determine the presence of mutations or even sequence the nucleic acid being analyzed or one can measure expression levels of a gene of interest. The method is one of parallel processing of many, even thousands, of probes at once and can tremendously increase the rate of analysis. Several papers have been published which use this technique. Some of these are Hacia et al., 1996; Shoemaker et al., 1996; Chee et al., 1996; Lockhart et al., 1996; DeRisi et al., 1996; Lipshutz et al., 1995. This method has already been used to screen people for mutations in the breast cancer gene BRCA1 (Hacia et al., 1996). This new technology has been reviewed in a news article in Chemical and Engineering News (Borman, 1996) and been the subject of an editorial (Nature Genetics, 1996). Also see Fodor (1997).

The most definitive test for mutations in a candidate locus is to directly compare genomic HPC1 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of HPC1 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the HPC1 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of HPC1 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type HPC1 gene. Alteration of wild-type HPC1 genes can also be detected by screening for alteration of wild-type HPC1 protein. For example, monoclonal antibodies immunoreactive with HPC1 can be used to screen a tissue. Lack of cognate antigen would indicate an HPC1 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant HPC1 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered HPC1 protein can be used to detect alteration of wild-type HPC1 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect HPC1 biochemical function. Finding a mutant HPC1 gene product indicates alteration of a wild-type HPC1 gene.

Mutant HPC1 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant HPC1 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the HPC1 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant HPC1 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which HPC1 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular HPC1 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the HPC1 gene on chromosome 1 in order to prime amplifying DNA synthesis of the HPC1 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the HPC1 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular HPC1 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from HPC1 sequences or sequences adjacent to HPC1, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the HPC1 open reading frame shown in SEQ ID NOs:1-52 (see Table 9), design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the HPC1 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type HPC1 gene do not have cancer which results from the HPC1 allele. However, mutations which interfere with the function of the HPC1 protein are involved in the pathogenesis of cancer. Thus, the presence of an altered (or a mutant) HPC1 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect an HPC1 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the HPC1 allele being analyzed and the sequence of the wild-type HPC1 allele. Mutant HPC1 alleles can be initially identified by any of the techniques described above. The mutant alleles are then

sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant HPC1 alleles can be initially identified by identifying mutant (altered) HPC1 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those which lead to an altered function of the HPC1 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions:

"**Amplification of Polynucleotides**" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. Also useful are strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA). These methods are well known and widely practiced in the art. See, e.g., U.S. Patents 4,683,195 and 4,683,202 and Innis *et al.*, 1990 (for PCR); and Wu *et al.*, 1989a (for LCR); U.S. Patents 5,270,184 and 5,455,166 (for SDA); Spargo et al., 1996 (for thermophilic SDA) and U.S. Patent 5,409,818, Fahy et al., 1991 and Compton, 1991 for 3SR and NASBA. Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the HPC1 region are preferably complementary to, and hybridize specifically to sequences in the HPC1 region or in regions that flank a target region therein. HPC1 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"**Analyte polynucleotide**" and "**analyte strand**" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"**Antibodies.**" The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the HPC1 polypeptides and fragments thereof or to polynucleotide sequences from the HPC1 region, particularly from the HPC1 locus or a portion thereof. The term "**antibody**" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier

molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the HPC1 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with HPC1 polypeptide or fragments thereof. See, Harlow and Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by *in vitro* or *in vivo* techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow and Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M^{-1} or preferably 10^{-9} to 10^{-10} M^{-1} or stronger will typically be made by standard procedures as described, e.g., in Harlow and Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse *et al.*, 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining,

either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Patent 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. It is well recognized by those of skill in the art that lengths shorter than 15, between 15 and 40, and greater than 40 bases may also be used. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A **"biological sample"** refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents.

As used herein, the terms **"diagnosing"** or **"prognosing,"** as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or **"substantially pure"**. An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"HPC1 Allele" refers to normal alleles of the HPC1 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian, colorectal and prostate cancer. Such predisposing alleles are also called **"HPC1 susceptibility alleles"**.

"HPC1 Locus", "HPC1 Gene", "HPC1 Nucleic Acids" or **"HPC1 Polynucleotide"** each refer to polynucleotides, all of which are in the HPC1 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal and prostate cancers. Mutations at the HPC1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the HPC1 region described *infra*. The HPC1 locus is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The HPC1 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes an HPC1 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural HPC1-encoding gene or one having substantial homology with a natural HPC1-encoding gene or a portion thereof.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates,

phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the HPC1 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7-15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with an HPC1-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available

from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank, National Institutes of Health.

"HPC1 Region" refers to a portion of human chromosome 1 bounded by the markers mM.GAAA.158;23.4 and mM.GA57e15.S16. This region contains the HPC1 locus, including the HPC1 gene.

As used herein, the terms **"HPC1 locus", "HPC1 allele"** and **"HPC1 region"** all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a **"portion"** of the *HPC1* locus or region or allele is defined as having a minimal size of at least about eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides. This definition includes all sizes in the range of 8-40 nucleotides as well as greater than 40 nucleotides. Thus, this definition includes nucleic acids of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400, 500 nucleotides, or nucleic acids having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72, 121, etc., nucleotides), or nucleic acids having more than 500 nucleotides. The present invention includes all novel nucleic acids having at least 8 nucleotides derived from any of SEQ ID NOs:1-52 and any combination of these sequences as described in further detail below, its complement or functionally equivalent nucleic acid sequences. The present invention does not include nucleic acids which exist in the prior art. That is, the present invention includes all nucleic acids having at least 8 nucleotides derived from any of SEQ ID NOs:1-52 and any combination of these sequences as described in further detail below with the proviso that it does not include nucleic acids existing in the prior art.

"HPC1 protein" or **"HPC1 polypeptide"** refers to a protein or polypeptide encoded by the HPC1 locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example,

glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native HPC1 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to HPC1-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the HPC1 protein(s).

An HPC1 polypeptide may be that derived from any of the exons described herein which may be in isolated and/or purified form, free or substantially free of material with which it is naturally associated. The polypeptide may, if produced by expression in a prokaryotic cell or produced synthetically, lack native post-translational processing, such as glycosylation. Alternatively, the present invention is also directed to polypeptides which are sequence variants, alleles or derivatives of an HPC1 polypeptide. Such polypeptides may have an amino acid sequence which differs from that derived form any of the exons described herein by one or more of addition, substitution, deletion or insertion of one or more amino acids. Preferred such polypeptides have HPC1 function.

Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, such as stability against proteolytic cleavage, without the loss of other functions or properties. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. Preferred substitutions are ones which are conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and tyrosine, phenylalanine.

Certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules or binding sites on proteins interacting with an HPC1 polypeptide. Since it is the interactive capacity and nature of

a protein which defines that protein's biological functional activity, certain amino acid substitutions can be made in a protein sequence, and its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydrophobic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, 1982). Alternatively, the substitution of like amino acids can be made effectively on the basis of hydrophilicity. The importance of hydrophilicity in conferring interactive biological function of a protein is generally understood in the art (U.S. Patent 4,554,101). The use of the hydrophobic index or hydrophilicity in designing polypeptides is further discussed in U.S. Patent 5,691,198.

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

The term **peptide mimetic** or **mimetic** is intended to refer to a substance which has the essential biological activity of an HPC1 polypeptide. A peptide mimetic may be a peptide-containing molecule that mimics elements of protein secondary structure (Johnson et al., 1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen, enzyme and substrate or scaffolding proteins. A peptide mimetic is designed to permit molecular interactions similar to the natural molecule. A mimetic may not be a peptide at all, but it will retain the essential biological activity of a natural HPC1 polypeptide.

"Probes". Polynucleotide polymorphisms associated with HPC1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under highly stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, high stringency conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely

complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. (It should be noted that throughout this disclosure, if it is simply stated that "stringent" conditions are used that is meant to be read as "high stringency" conditions are used.) Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of an HPC1 susceptibility allele.

Probes for HPC1 alleles may be derived from the sequences of the HPC1 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the HPC1 region, and which allow specific hybridization to the HPC1 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid will be relatively stable under even highly stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding HPC1 are preferred as probes. Thus, this definition includes probes of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400 or 500 nucleotides or probes having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72,

121, etc., nucleotides), or probes having more than 500 nucleotides. The probes may also be used to determine whether mRNA encoding HPC1 is present in a cell or tissue. The present invention includes all novel probes having at least 8 nucleotides derived from any of SEQ ID NOs:1-42 and any combination of these sequences as described in further detail below, its complement or functionally equivalent nucleic acid sequences. The present invention does not include probes which exist in the prior art. That is, the present invention includes all probes having at least 8 nucleotides derived from any of SEQ ID NOs:1-52 and any combination of these sequences as described in further detail below with the proviso that they do not include probes existing in the prior art.

Similar considerations and nucleotide lengths are also applicable to primers which may be used for the amplification of all or part of the *HPC1* gene. Thus, a definition for primers includes primers of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400, 500 nucleotides, or primers having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72, 121, etc. nucleotides), or primers having more than 500 nucleotides, or any number of nucleotides between 500 and 9000. The primers may also be used to determine whether mRNA encoding *HPC1* is present in a cell or tissue. The present invention includes all novel primers having at least 8 nucleotides derived from the *HPC1* locus for amplifying the *HPC1* gene, its complement or functionally equivalent nucleic acid sequences. The present invention does not include primers which exist in the prior art. That is, the present invention includes all primers having at least 8 nucleotides with the proviso that it does not include primers existing in the prior art.

"Protein modifications or fragments" are provided by the present invention for HPC1 polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability

requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of HPC1 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the HPC1 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides will serve as highly efficient competitors for specific binding. Production of antibodies specific for HPC1 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising HPC1 polypeptides and fragments. Homologous polypeptides may be fusions between two or more HPC1 polypeptide sequences or between the sequences of HPC1 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial b-galactosidase, trpE, protein A, b-lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See Godowski *et al.*, 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the HPC1 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding HPC1, and are well known in the art. For example, such polypeptides

may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "**isolated**", "**substantially pure**", and "**substantially homogeneous**" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

An HPC1 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"**Recombinant nucleic acid**" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"**Regulatory sequences**" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"**Substantial homology or similarity**". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

To determine homology between two different nucleic acids, the percent homology is to be determined using the BLASTN program "BLAST 2 sequences". This program is available for public use from the National Center for Biotechnology Information (NCBI) over the Internet (http://www.ncbi.nlm.nih.gov/gorf/bl2.html) (Altschul et al., 1997). The parameters to be used are whatever combination of the following yields the highest calculated percent homology (as calculated below) with the default parameters shown in parentheses:

Program - blastn
Matrix - 0 BLOSUM62
Reward for a match - 0 or 1 (1)
Penalty for a mismatch - 0, -1, -2 or -3 (-2)
Open gap penalty - 0, 1, 2, 3, 4 or 5 (5)
Extension gap penalty - 0 or 1 (1)
Gap x_dropoff - 0 or 50 (50)
Expect - 10

Along with a variety of other results, this program shows a percent identity across the complete strands or across regions of the two nucleic acids being matched. The program shows as part of the results an alignment and identity of the two strands being compared. If the strands are of equal length then the identity will be calculated across the complete length of the nucleic acids. If the strands are of unequal lengths, then the length of the shorter nucleic acid is to be used. If the nucleic acids are quite similar across a portion of their sequences but different across the rest of their sequences, the blastn program "BLAST 2 Sequences" will show an identity across only the similar portions, and these portions are reported individually. For

purposes of determining homology herein, the percent homology refers to the shorter of the two sequences being compared. If any one region is shown in different alignments with differing percent identities, the alignments which yield the greatest homology are to be used. The averaging is to be performed as in this example of SEQ ID NOs:53 and 54.

5'-ACCGTAGCTACGTACGTATATAGAAAGGGCGCGATCGTCGTCGCGTATG ACGACTTAGCATGC-3' (SEQ ID NO:53)

5'-ACCGGTAGCTACGTACGTTATTTAGAAAGGGGTGTGTGTGTGTGTGTAA ACCGGGGTTTTCGGGATCGTCCGTCGCGTATGACGACTTAGCCATGCACGGTATAT CGTATTAGGACTAGCGATTGACTAG-3' (SEQ ID NO:54)

The program "BLAST 2 Sequences" shows differing alignments of these two nucleic acids depending upon the parameters which are selected. As examples, four sets of parameters were selected for comparing SEQ ID NOs:53 and 54 (gap x_dropoff was 50 for all cases), with the results shown in Table A. It is to be noted that none of the sets of parameters selected as shown in Table A is necessarily the best set of parameters for comparing these sequences. The percent homology is calculated by multiplying for each region showing identity the fraction of bases of the shorter strand within a region times the percent identity for that region and adding all of these together. For example, using the first set of parameters shown in Table A, SEQ ID NO:53 is the short sequence (63 bases), and two regions of identity are shown, the first encompassing bases 4-29 (26 bases) of SEQ ID NO:53 with 92% identity to SEQ ID NO:54 and the second encompassing bases 39-59 (21 bases) of SEQ ID NO:53 with 100% identity to SEQ ID NO:54. Bases 1-3, 30-38 and 60-63 (16 bases) are not shown as having any identity with SEQ ID NO:54. Percent homology is calculated as: (26/63)(92) + (21/63)(100) + (16/63)(0) = 71.3% homology. The percents of homology calculated using each of the four sets of parameters shown are listed in Table A. Several other combinations of parameters are possible, but they are not listed for the sake of brevity. It is seen that each set of parameters resulted in a different calculated percent homology. Because the result yielding the highest percent homology is to be used, based solely on these four sets of parameters one would state that SEQ ID NOs:53 and 54 have 87.1% homology. Again it is to be noted that use of other parameters may show an even higher homology for SEQ ID NOs:53 and 54, but for brevity not all the possible results are shown.

TABLE A

Parameter Values				Regions of identity (%)		Homology
Match	Mismatch	Open Gap	Extension Gap			
1	-2	5	1	4-29 of 5 and 5-31 of 6 (92%)	39-59 of 5 and 71-91 of 6 (100%)	71.3
1	-2	2	1	4-29 of 5 and 5-31 of 6 (92%)	33-63 of 5 and 64-96 of 6 (93%)	83.7
1	-1	5	1	----------	30-59 of 5 and 61-91 of 6 (93%)	44.3
1	-1	2	1	4-29 of 5 and 5-31 of 6 (92%)	30-63 of 5 and 61-96 of 6 (91%)	87.1

Identity means the degree of sequence relatedness between two polypeptide or two polynucleotides sequences as determined by the identity of the match between two strings of such sequences. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Methods commonly employed to determine identity between two sequences include, but are not limited to those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994,

and Carillo, H., and Lipman, D. (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux et al. (1984), BLASTP, BLASTN, FASTA (Altschul et al. (1990)).

Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms **"substantial homology"** or **"substantial identity"**, when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30%

identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measures of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

"**Substantially similar function**" refers to the function of a modified nucleic acid or a modified protein, with reference to the wild-type HPC1 nucleic acid or wild-type HPC1 polypeptide. The modified polypeptide will be substantially homologous to the wild-type HPC1 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type HPC1 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type HPC1 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type HPC1 gene function produces the modified protein described above.

A polypeptide "**fragment**," "**portion**" or "**segment**" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"**Target region**" refers to a region of the nucleic acid which is amplified and/or detected. The term "**target sequence**" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis *et al.*, 1982; Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; Glover, 1985; Anand, 1992; Guthrie and Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 1, is provided, e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage and Carruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding

segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native HPC1 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.* 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with HPC1 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992; see also, e.g., Metzger *et al.*, 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England BioLabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman *et al.*, EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers *et al.*, 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1983). See also, e.g., U.S. Patent Nos. 5,691,198; 5,735,500; 5,747,469 and 5,436,146.

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo *et al.*, 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook *et al.*, 1989 and Ausubel *et al.*, 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, *inter alia*, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the HPC1 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is *per se* well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines. An example of a commonly used insect cell line is SF9. However, it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of HPC1 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the HPC1 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the HPC1 locus or other sequences from the HPC1 region (particularly those flanking the HPC1 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with HPC1 transcription and/or translation and/or replication.

The probes and primers based on the HPC1 gene sequences disclosed herein are used to identify homologous HPC1 gene sequences and proteins in other species. These HPC1 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of an HPC1 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of HPC1. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of HPC1. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant HPC1 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based

strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 1. Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis *et al.*, 1982 and Sambrook *et al.*, 1989. Under certain circumstances, the formation of higher order hybrids, such as triplexes, quadraplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding

with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews and Kricka, 1988; Landegren *et al.*, 1988; Mittlin, 1989; U.S. Patent 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3-10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski *et al.*, 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding HPC1. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing or potentially predisposing mutations summarized in Tables 9 and 10 of this patent application.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin *et al.*, 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the

biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby *et al.*, 1977 and Nguyen *et al.*, 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting HPC1. Thus, in one example to detect the presence of HPC1 in a cell sample, more than one probe complementary to HPC1 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the HPC1 gene sequence in a patient, more than one probe complementary to HPC1 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in HPC1. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to prostate cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations identified in Tables 9 and 10 and those that have the HPC1 regions corresponding to SEQ ID NOs:1-52 both 5' and 3' to the mutation site.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type HPC1 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of, HPC1 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 12 and 13. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate HPC1 proteins from solution as well as react with HPC1 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect HPC1 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting HPC1 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David *et*

al. in U.S. Patent Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 15.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the HPC1 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The HPC1 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between an HPC1 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between an HPC1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with an HPC1 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the HPC1 polypeptide or fragment, or (ii) for the presence of a complex between the HPC1 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the HPC1 polypeptide or fragment is typically labeled. Free HPC1 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to HPC1 or its interference with HPC1:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the HPC1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on September 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with HPC1 polypeptide and washed. Bound HPC1 polypeptide is then detected by methods well known in the art.

Purified HPC1 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the HPC1 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the HPC1 polypeptide compete with a test compound for binding to the HPC1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the HPC1 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional HPC1 gene. These host cell lines or cells are defective at the HPC1 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of HPC1 defective cells.

Briefly, a method of screening for a substance which modulates activity of a polypeptide may include contacting one or more test substances with the polypeptide in a suitable reaction medium, testing the activity of the treated polypeptide and comparing that activity with the activity of the polypeptide in comparable reaction medium untreated with the test substance or substances. A difference in activity between the treated and untreated polypeptides is indicative of a modulating effect of the relevant test substance or substances.

Prior to or as well as being screened for modulation of activity, test substances may be screened for ability to interact with the polypeptide, e.g., in a yeast two-hybrid system (e.g., Bartel et al., 1993; Fields and Song, 1989; Chevray and Nathans, 1992; Lee et al., 1995). This system may be used as a coarse screen prior to testing a substance for actual ability to modulate activity of the polypeptide. Alternatively, the screen could be used to screen test substances for binding to an HPC1 specific binding partner, or to find mimetics of an HPC1 polypeptide.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional

structure of a protein of interest (e.g., HPC1 polypeptide) or, for example, of the HPC1-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson *et al.*, 1990). In addition, peptides (e.g., HPC1 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved HPC1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of HPC1 polypeptide activity. By virtue of the availability of cloned HPC1 sequences, sufficient amounts of the HPC1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the HPC1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Following identification of a substance which modulates or affects polypeptide activity, the substance may be investigated further. Furthermore, it may be manufactured and/or used in preparation, i.e., manufacture or formulation, or a composition such as a medicament, pharmaceutical composition or drug. These may be administered to individuals.

Thus, the present invention extends in various aspects not only to a substance identified using a nucleic acid molecule as a modulator of polypeptide activity, in accordance with what is disclosed herein, but also a pharmaceutical composition, medicament, drug or other composition comprising such a substance, a method comprising administration of such a composition comprising such a substance, a method comprising administration of such a composition to a

patient, e.g., for treatment of prostate cancer, use of such a substance in the manufacture of a composition for administration, e.g., for treatment of prostate cancer, and a method of making a pharmaceutical composition comprising admixing such a substance with a pharmaceutically acceptable excipient, vehicle or carrier, and optionally other ingredients.

A substance identified as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide "small molecules" are often preferred for many *in vivo* pharmaceutical uses. Accordingly, a mimetic or mimic of the substance (particularly if a peptide) may be designed for pharmaceutical use.

The designing of mimetics to a known pharmaceutically active compound is a known approach to the development of pharmaceuticals based on a "lead" compound. This might be desirable where the active compound is difficult or expensive to synthesize or where it is unsuitable for a particular method of administration, e.g., pure peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large numbers of molecules for a target property.

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. First, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, e.g., by substituting each residue in turn. Alanine scans of peptide are commonly used to refine such peptide motifs. These parts or residues constituting the active region of the compound are known as its "pharmacophore".

Once the pharmacophore has been found, its structure is modeled according to its physical properties, e.g., stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g., spectroscopic techniques, x-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modeling process.

In a variant of this approach, the three-dimensional structure of the ligand and its binding partner are modeled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design of the mimetic.

A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted onto it can conveniently be selected so that the mimetic is easy to synthesize, is likely to be pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide-based, further stability can be achieved by cyclizing the peptide, increasing its rigidity. The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimization or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type HPC1 function to a cell which carries mutant HPC1 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type HPC1 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant HPC1 allele, the gene fragment should encode a part of the HPC1 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type HPC1 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant HPC1 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the HPC1 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate coprecipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type HPC1 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the HPC1 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of HPC1 polypeptide is absent or

diminished compared to normal cells. It may also be useful to increase the level of expression of a given HPC1 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of HPC1 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the HPC1 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Patent 5,252,479 and PCT published application WO 93/07282 and U.S. Patent Nos. 5,691,198; 5,747,469; 5,436,146 and 5,753,500.. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak *et al.*, 1992), adenovirus (Berkner, 1992; Berkner *et al.*, 1988; Gorziglia and Kapikian, 1992; Quantin *et al.*, 1992; Rosenfeld *et al.*, 1992; Wilkinson *et al.*, 1992; Stratford-Perricaudet *et al.*, 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi *et al.*, 1990; Russell and Hirata, 1998), herpes viruses including HSV and EBV (Margolskee, 1992; Johnson *et al.*, 1992; Fink *et al.*, 1992; Breakfield and Geller, 1987; Freese *et al.*, 1990; Fink et al., 1996), lentiviruses (Naldini et al., 1996), Sindbis and Semliki Forest virus (Berglund et al., 1993), and retroviruses of avian (Bandyopadhyay and Temin, 1984; Petropoulos *et al.*, 1992), murine (Miller, 1992; Miller *et al.*, 1985; Sorge *et al.*, 1984; Mann and Baltimore, 1985; Miller *et al.*, 1988), and human origin (Shimada *et al.*, 1991; Helseth *et al.*, 1990; Page *et al.*, 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer *et al.*, 1980); mechanical techniques, for example microinjection (Anderson *et al.*, 1980; Gordon *et al.*, 1980; Brinster *et al.*, 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via

liposomes (Felgner *et al.*, 1987; Wang and Huang, 1989; Kaneda et al, 1989; Stewart *et al.*, 1992; Nabel *et al.*, 1990; Lim *et al.*, 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff *et al.*, 1990; Wu *et al.*, 1991; Zenke *et al.*, 1990; Wu *et al.*, 1989b; Wolff *et al.*, 1991; Wagner *et al.*, 1990; Wagner *et al.*, 1991; Cotten *et al.*, 1990; Curiel *et al.*, 1991a; Curiel *et al.*, 1991b). Viral-mediated gene transfer can be combined with direct *in vivo* gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver *et al.*, 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged. For other techniques for the delivery of adenovirus based vectors see Schneider et al. (1998) and U.S. Patent Nos. 5,691,198; 5,747,469; 5,436,146 and 5,753,500.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumor deposits, for example, following direct *in situ* administration (Nabel, 1992).

Expression vectors in the context of gene therapy are meant to include those constructs containing sequences sufficient to express a polynucleotide that has been cloned therein. In viral expression vectors, the construct contains viral sequences sufficient to support packaging of the construct. If the polynucleotide encodes *HPC1*, expression will produce HPC1. If the polynucleotide encodes an antisense polynucleotide or a ribozyme, expression will produce the antisense polynucleotide or ribozyme. Thus in this context, expression does not require that a protein product be synthesized. In addition to the polynucleotide cloned into the expression vector, the vector also contains a promoter functional in eukaryotic cells. The cloned polynucleotide sequence is under control of this promoter. Suitable eukaryotic promoters include those described above. The expression vector may also include sequences, such as selectable markers and other sequences described herein.

Gene transfer techniques which target DNA directly to prostate tissues, e.g., epithelial cells of the prostate, are preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry an HPC1 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal HPC1 allele. In this step, the treated individuals have reduced risk of prostate cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the prostate of a full term pregnancy.

Methods of Use: Peptide Therapy

Peptides which have HPC1 activity can be supplied to cells which carry mutant or missing HPC1 alleles. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, HPC1 polypeptide can be extracted from HPC1-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize HPC1 protein. Any of such techniques can provide the preparation of the present invention which comprises the HPC1 protein. Preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or *in vitro*.

Active HPC1 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the HPC1 gene product may be sufficient to affect tumor growth. Supply of molecules with HPC1 activity should lead to partial reversal of the

neoplastic state. Other molecules with HPC1 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant HPC1 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with HPC1 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the HPC1 allele, as described above. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant HPC1 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous *HPC1* gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecchi, 1989; Valancius and Smithies, 1991; Hasty *et al.*, 1991; Shinkai *et al.*, 1992; Mombaerts *et al.*, 1992; Philpott *et al.*, 1992; Snouwaert *et al.*, 1992; Donehower *et al.*, 1992) to produce knockout or transplacement animals. A transplacement is similar to a knockout because the endogenous gene is replaced, but in the case of a transplacement the replacement is by another version of the same gene. After test substances have been administered to the animals, the phenotype must be assessed. If the test substance prevents or suppresses the disease, then the test substance is a candidate therapeutic agent for the treatment of disease. These animal models provide an extremely important testing vehicle for potential therapeutic products.

In one embodiment of the invention, transgenic animals are produced which contain a functional transgene encoding a functional HPC1 polypeptide or variants thereof. Transgenic animals expressing *HPC1* transgenes, recombinant cell lines derived from such animals and transgenic embryos may be useful in methods for screening for and identifying agents that

induce or repress function of HPC1. Transgenic animals of the present invention also can be used as models for studying indications such as cancer.

In one embodiment of the invention, a *HPC1* transgene is introduced into a non-human host to produce a transgenic animal expressing a human or murine *HPC1* gene. The transgenic animal is produced by the integration of the transgene into the genome in a manner that permits the expression of the transgene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Patent No. 4,873,191; which is incorporated herein by reference), Brinster *et al.* 1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A Laboratory Manual" 2nd edition (eds., Hogan, Beddington, Costantimi and Long, Cold Spring Harbor Laboratory Press, 1994; which is incorporated herein by reference in its entirety).

It may be desirable to replace the endogenous *HPC1* by homologous recombination between the transgene and the endogenous gene; or the endogenous gene may be eliminated by deletion as in the preparation of "knock-out" animals. Typically, a *HPC1* gene flanked by genomic sequences is transferred by microinjection into a fertilized egg. The microinjected eggs are implanted into a host female, and the progeny are screened for the expression of the transgene. Transgenic animals may be produced from the fertilized eggs from a number of animals including, but not limited to reptiles, amphibians, birds, mammals, and fish. Within a particularly preferred embodiment, transgenic mice are generated which overexpress HPC1 or express a mutant form of the polypeptide. Alternatively, the absence of a *HPC1* in "knock-out" mice permits the study of the effects that loss of HPC1 protein has on a cell *in vivo*. Knock-out mice also provide a model for the development of HPC1-related cancers.

Methods for producing knockout animals are generally described by Shastry (1995, 1998) and Osterrieder and Wolf (1998). The production of conditional knockout animals, in which the gene is active until knocked out at the desired time is generally described by Feil et al. (1996), Gagneten et al. (1997) and Lobe and Nagy (1998). Each of these references is incorporated herein by reference.

As noted above, transgenic animals and cell lines derived from such animals may find use in certain testing experiments. In this regard, transgenic animals and cell lines capable of expressing wild-type or mutant HPC1 may be exposed to test substances. These test substances can be screened for the ability to reduce overepression of wild-type *HPC1* or impair the expression or function of mutant HPC1.

Pharmaceutical Compositions and Routes of Administration

The HPC1 polypeptides, antibodies, peptides and nucleic acids of the present invention can be formulated in pharmaceutical compositions, which are prepared according to conventional pharmaceutical compounding techniques. See, for example, Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA). The composition may contain the active agent or pharmaceutically acceptable salts of the active agent. These compositions may comprise, in addition to one of the active substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The carrier may take a wide variety of forms depending on the form of preparation desired for administration, e.g., intravenous, oral, intrathecal, epineural or parenteral.

For oral administration, the compounds can be formulated into solid or liquid preparations such as capsules, pills, tablets, lozenges, melts, powders, suspensions or emulsions. In preparing the compositions in oral dosage form, any of the usual pharmaceutical media may be employed, such as, for example, water, glycols, oils, alcohols, flavoring agents, preservatives, coloring agents, suspending agents, and the like in the case of oral liquid preparations (such as, for example, suspensions, elixirs and solutions); or carriers such as starches, sugars, diluents, granulating agents, lubricants, binders, disintegrating agents and the like in the case of oral solid preparations (such as, for example, powders, capsules and tablets). Because of their ease in administration, tablets and capsules represent the most advantageous oral dosage unit form, in which case solid pharmaceutical carriers are obviously employed. If desired, tablets may be sugar-coated or enteric-coated by standard techniques. The active agent can be encapsulated to make it stable to passage through the gastrointestinal tract while at the same time allowing for passage across the blood brain barrier. See for example, WO 96/11698.

For parenteral administration, the compound may be dissolved in a pharmaceutical carrier and administered as either a solution or a suspension. Illustrative of suitable carriers are water, saline, dextrose solutions, fructose solutions, ethanol, or oils of animal, vegetative or synthetic origin. The carrier may also contain other ingredients, for example, preservatives, suspending agents, solubilizing agents, buffers and the like. When the compounds are being administered intrathecally, they may also be dissolved in cerebrospinal fluid.

The active agent is preferably administered in a therapeutically effective amount. The actual amount administered, and the rate and time-course of administration, will depend on the

nature and severity of the condition being treated. Prescription of treatment, e.g. decisions on dosage, timing, etc., is within the responsibility of general practitioners or specialists, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of techniques and protocols can be found in *Remington's Pharmaceutical Sciences*.

Alternatively, targeting therapies may be used to deliver the active agent more specifically to certain types of cell, by the use of targeting systems such as antibodies or cell specific ligands. Targeting may be desirable for a variety of reasons, e.g. if the agent is unacceptably toxic, or if it would otherwise require too high a dosage, or if it would not otherwise be able to enter the target cells.

Instead of administering these agents directly, they could be produced in the target cell, e.g. in a viral vector such as described above or in a cell based delivery system such as described in U.S. Patent No. 5,550,050 and published PCT application Nos. WO 92/19195, WO 94/25503, WO 95/01203, WO 95/05452, WO 96/02286, WO 96/02646, WO 96/40871, WO 96/40959 and WO 97/12635, designed for implantation in a patient. The vector could be targeted to the specific cells to be treated, or it could contain regulatory elements which are more tissue specific to the target cells. The cell based delivery system is designed to be implanted in a patient's body at the desired target site and contains a coding sequence for the active agent. Alternatively, the agent could be administered in a precursor form for conversion to the active form by an activating agent produced in, or targeted to, the cells to be treated. See for example, EP 425,731A and WO 90/07936.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a Chromosome 1-Linked Prostate Cancer Susceptibility Locus

Extensive cancer prone kindreds were ascertained from a defined population providing a large set of extended kindreds with multiple cases of prostate cancer and many relatives

available to study. The large number of meioses present in these large kindreds provided the power to detect whether the HPC1 locus was segregating, and increased the opportunity for informative recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the HPC1 region, and greatly facilitated the reduction of the HPC1 region to a manageable size, which permits identification of the HPC1 locus itself.

Each kindred was extended through all available connecting relatives, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional prostate cancer cases and individuals with cancer at other sites of interest (e.g., bladder) who also appeared in the kindreds were identified through the tumor registry linked files. All prostate cancers reported in the kindred which were not confirmed in the Utah Cancer Registry were verified. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited to participate by providing a blood sample from which DNA was extracted. We also sampled spouses, siblings, and offspring of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Each of the Utah pedigrees studied represents the descendants of a single founder for whom a significant excess of prostate cancer cases was observed among all descendants. Since all affected descendants are studied, the resulting kindreds represent a collection of both closely and distantly related prostate cancer cases. The criteria for selection of kindreds to analyze for HPC1 linkage were: 1) genotypes available, or inferable, for 6 or more prostate cancer cases, and 2) at least 3 genotyped cases within a second degree of relationship to another genotyped case.

The Utah kindreds are 5 - 7 generations deep, and contain between 8 and 29 prostate cancer cases. They are all Caucasian of Northern European ancestry. The median age-of-onset for each kindred ranged from 64 to 76, similar to that estimated for the general population. Five percent of cases were diagnosed before age 55.

For each kindred analyzed, the number of prostate cancer cases, the median age and range of age-of-onset, and the number of cases and family members sampled and included in this analysis are detailed in Table 1. The kindreds labeled A-E in Table 1 are the kindreds used for the data which are shown in Table 3.

Table 1

The 29 Utah Kindreds

Kindred	Total Cases*	Age-of-Onset Range	Age-of-Onset Median	Case Samples+	Total Samples^
1	9	57-80	73	4	12
2	8	44-91	66	7	24
3	8	56-79	76	4	20
4	18	56-88	74	8	40
5	19	57-88	73	7	49
6	7	60-88	73	14	81
8	16	46-84	76	1	39
9	13	61-82	76	5	51
10 (A)	15	55-88	71	7	31
11	16	60-82	70	8	34
12 (B)	14	56-85	73	9	41
13	14	50-88	68	6	29
14	10	51-85	68	4	34
15	11	45-85	68	4	30
16	17	44-84	71	11	41
17 (C)	11	44-86	66	7	87
18	11	47-81	70	5	22
19	14	54-86	72	5	21
20	8	62-81	71	5	12
21	12	45-83	71	4	21
22	11	58-91	76	7	25
23	8	51-84	64	3	16
24	21	54-87	65	15	41
25 (D)	8	56-78	68	4	34
26	8	60-77	70	3	29
27	11	62-87	67	7	37
28 (E)	10	53-86	67	5	26
29	11	45-86	0	6	14
Totals	368			190	959

*Total affected individuals in the genotyped portion of the kindred.
+Total affected individuals genotyped for the three markers
^Total individuals genotyped for the three markers (includes affected samples).

EXAMPLE 2

Selection of Kindreds Which are Linked to Chromosome 1 and Localization of HPC1 to the Interval mM.GAAA158;23.4 - mM.GA57e15.S16

Nuclear pellets were extracted from 16 ml of ACD blood, and DNA extracted with phenol and chloroform, precipitated with ethanol, and resuspended in Tris-EDTA. The markers

used for genotyping were short tandem repeat (STR) loci at 1q24-25 which flanked the most likely HPC1 location as indicated in Smith et al. (1996). The order of markers designated by D followed by 1S followed by a sequential number and approximate intervals between them (in centiMorgans) is:

D1S2883 -10.8 centiMorgans - D1S254 - 11.5 centiMorgans - D1S412.

The most likely location as suggested in Smith et al. (1996) is at D1S254.

Amplification of 20 ng genomic DNA was performed according to standard PCR procedures, with minor modifications to optimize product clarity, in a total reaction mix of 10 ml. Radiolabeled PCR products were electrophoresed on standard 6% polyacrylamide denaturing sequencing gels. Gels were then dried and autoradiographed. A total of over 200 prostate cancer cases and approximately 800 of their relatives were genotyped for the markers.

In the kindreds which showed evidence of segregation, up to an additional 35 markers were used to identify and confirm segregation of multiple linked markers (haplotypes). These markers were spaced throughout the 28.7 cM region between D1S452 (proximal to D1S2883) and D1S422 (distal to D1S412), a region which flanks the three originally typed markers by 3.6 cM distally, and 2.8 cM proximally.

Two-point linkage analysis was performed with the package LINKAGE (Lathrop et al., 1984; 1985) using the FASTLINK implementation (Cottingham et al., 1993; Schaffer et al., 1994). The statistical analysis for the inheritance of susceptibility to prostate cancer used the model described in Smith et al. (1996). This model assumed a rare autosomal dominant susceptibility locus and allowed for a 15% sporadic rate of prostate cancer. Marker allele frequencies were estimated from unrelated individuals present in the kindreds.

Linkage in the presence of heterogeneity was assessed by the admixture test (A-test) of Ott (1986). HOMOG, which postulates two family types, linked and unlinked, was used. Multipoint linkage analysis was performed using VITESSE (O'Connell et al., 1995). The size of the pedigrees and the lack of genotyping of the higher generations due to the late age-of-onset, made more-than-three-point analyses impossible. The multipoint results in Figure 3 represent a walking three-point analysis, with the disease phenotype placed between each pair of adjacent markers in all intervals but the exterior ones, in which the two closest markers were used.

The two-point Lod scores for the 29 kindreds combined were highly negative at the 3 markers examined (Table 2), suggesting an overall lack of evidence for this susceptibility locus across all kindreds. Heterogeneity analysis of the three loci showed weak, non-significant

evidence for one locus, explaining 5% of the pedigrees. The positive Lod score observed for D1S254 in analysis of heterogeneity, as well as the low estimate of alpha reported in Smith et al. (1996) suggested that there might be a subset of linked pedigrees within our data set. We examined three marker haplotypes in each kindred for evidence of a shared region among affecteds. For those kindreds which suggested such segregation, we genotyped samples for up to an additional 35 markers. In Table 2 these kindreds and their Lod scores for the 3 markers are shown.

Multipoint linkage results are depicted in Figure 3. This analysis resulted in a maximum heterogeneity Lod score of +1.20 at D1S254 with an estimate that 5% of kindreds were linked. Multipoint heterogeneity analysis in the most likely interval excluded linkage (Lod scores less than -2.00) for alpha greater than 0.33.

Cancers of sites other than prostate would also be expected to occur in individuals in these kindreds. Some individuals hypothesized to be sharing the segregating chromosome 1 haplotype were affected with cancer at another site. These included stomach cancers at ages 56, 68 and 82, ovarian cancer at age 32, and breast cancer at age 49 in kindred 17; a colon cancer at age 87, in kindred 12; and a breast cancer at age 72 and colon cancer at age 79 in kindred 25. Lod scores for linkage for a phenotype of cancer of any site did not differ significantly from those for prostate alone, although most individuals with cancer of another site were not included in the sampling.

Table 2

Total Lod scores and heterogeneity Lod scores for 29 Utah high-risk prostate cancer kindreds

	Lod Scores				
	r = 0.00	0.01	0.10	0.20	0.30
Marker					
D1S2883	-40.41	-33.31	-11.41	-3.70	-0.89
D1S254	-27.18	-21.90	- 6.47	-1.52	-0.08
D1S412	-46.31	-37.49	-12.45	-3.96	-1.00

	Heterogeneity	
Marker	Lod (r)	alpha
D1S2883	0.004 (.3)	0.10
D1S254	0.482 (.0)	0.05
D1S412	0.004 (.3)	0.05

Table 3

Maximum Lod scores for the 5 Utah kindreds with evidence of segregation of the three-marker haplotype

	Maximum Lod (r)		
Kindred	D1S2883	D1S254	D1S412
10	0.64(.0)	0.00(.5)	0.00(.5)
12	0.28(.2)	0.02(.3)	0.00(.5)
17	0.43(.0)	2.04(.0)	0.39(.1)
25	0.42(.0)	0.27(.0)	0.13(.2)
28	0.05(.0)	0.31(.1)	0.12(.1)

Maximum Lod score =0.00(.5) indicates no evidence for linkage.

Analysis model used for Tables 2 and 3:

Disease gene frequency: 0.003

Unaffecteds age < 75 years:

unknown phenotype

Unaffecteds age >= 75 years:

non carrier genotype disease penetrance = 0.16

carrier genotype disease penetrance = 0.63

Affecteds:

non carrier genotype disease penetrance = 0.00053

carrier genotype disease penetrance = 0.50

EXAMPLE 3

Contig Assembly

Genomic clone contig assembly in the HPC1 region started from a publicly available integrated map of chromosome 1, the WICGR Chr 1 map of Nov. 19, 1996. YACs located in the interval between D1S202 and D1S238 were ordered from Genome Systems (Figure 1). Primer pairs for the markers located in the interval between D1S202 and D1S238 were synthesized and used to screen a BAC library at Myriad. Markers that were negative on that BAC library were used to screen the BAC and PAC libraries at Genome Systems. DNA preps were prepared from the BACs and PACs that contained these markers. End sequences were obtained by dye terminator sequencing with vector primers on ABI 377 sequencers. Primer pairs defining BAC or PAC end markers were designed from these sequences. These new markers were checked against the YACs to make sure that they mapped within the interval. If the map data were ambiguous, the markers were also checked against a radiation hybrid panel. These new markers were checked against the already identified BACs/PACs to determine the positions of these clones relative to each other. The outside markers from each clone contig were used to screen the Myriad BAC library; those that were negative on that BAC library were used to screen the BAC and PAC libraries at Genome Systems. Repeated cycles of library screening and marker development allowed us to build a BAC/PAC contig that spanned the minimal recombinant interval.

As shown in our physical map of the HPC1 locus (Figure 1), a 15 clone BAC/PAC contig spans the interval between D1S202 and D1S238. Based on the genetic data described in detail above, the HPC1 locus must lie in the interval between the marker mM.GAAA158j23.4 and mM.GA57e15.S6. This interval is spanned by a 10 clone BAC/PAC contig. Based on the sizes and map positions of the YACs in the region, the sizes of these BACs and PACs in the contig and extensive sequencing of those BACs and PACs, we estimate the size of the minimal genetically defined interval containing HPC1 to be 750 kb.

EXAMPLE 4

Genomic sequencing

Two different types of genomic sequencing sublibraries were prepared from BAC or PAC clones in the candidate region.

Random-Sheared Sequencing Sub-libraries BAC or PAC DNA was sheared by sonication. To generate blunt-ended fragments, the sonicated DNA was incubated with mung-bean nuclease (Pharmacia Biotech) followed by treatment with a Pfu polishing kit (Stratagene). The DNA fragments were size fractionated on a 0.8% TAE agarose gel, and fragments in the size range of 1.0 - 1.6 kb were excised under longwave (365 nm) ultraviolet light. The excised gel slice was rotated 180 degrees relative to the original direction of electrophoresis and then placed into a new gel tray containing 1.0% GTG-Seaplaque low-melting temperature agarose (FMC corporation) before the gel solidified. Electrophoresis was repeated for the same time and voltage as the first run, resulting in a concentration of the DNA fragments in a small volume of agarose, and the gel slice containing the DNA fragments was once again excised from the gel. The DNA fragments were purified from the agarose by incubating the gel slice with beta-agarose (New England Biolabs), followed by removal of the agarose monomers using disposable microconcentrators (Amicon) that employ a 50,000 Daltons molecular weight cutoff filter. DNA fragments were ligated into the Hinc II site of the plasmid pMYG2, a pBluescript (Stratagene) derivative where the polylinker has been replaced by the pMYG2 polylinker. The vector was prepared by digestion with Hinc II followed by dephosphorylation with calf alkaline phosphatase (Boehringer Mannheim).

Table 4

Cloning Sites in pMYG1 and pMYG2

Name	Sequence	Sequence ID#
pMYG2 polylinker	ATGACCATAGTCGACCTGGCCGTCGTT	55
pMYG1 polylinker	ATGACCATAGTCGACGGATCCGTCGACCTG GCCGTCGTT	56

Ligated products were transformed into DH5-alpha *E. coli* competent cells (Life Technologies, Inc.) and plated on LB plates containing ampicillin, IPTG, and Bluo-gal (Sigma; Life Technologies, Inc.) . White colonies were used to inoculate individual wells of 1 ml 96-well microtiter plates (Beckman) containing 200 microliters of LB media supplemented with ampicillin at 50 micrograms per milliliter. The plates were incubated for 16-20 hours in a shaking incubator at 37 degrees Celsius. After incubation, 20 microliters of dimethyl sulfoxide was added to each well and the plates stored frozen. The inserts of random-sheared clones were amplified from *E. coli* cultures by PCR with vector primers, and the PCR products were

sequenced with M13 forward or reverse fluorescent energy transfer (FET) dye-labeled primers on ABI 377 sequencers.

Sau 3A Sequencing Sub-libraries: BAC or PAC DNA was partially digested with the restriction enzyme Sau 3A, and fragments in the size range of 5-8 kb were size fractionated and recovered from the agarose gel as described above for random-sheared fragments. Sau 3A fragments were ligated into the Bam HI site of pMYG1, a pBluescript (Stratagene) derivative where the polylinker has been replaced by the pMYG1 polylinker. The vector was prepared by digestion with Bam HI and dephosphorylation with shrimp alkaline phosphatase (Amersham). The ligated products were transformed and plated as described above for random-sheared clones.

To identify clones containing inserts in the size range of 5-8 kb, bacterial colonies were screened using a plasmid preparation procedure that has been adapted for use in a 96-well format. White colonies were picked into individual wells of 2 ml 96-well plates (Continental Laboratory Products) containing 1 ml LB media supplemented with 200 micrograms per milliliter ampicillin. The plates were incubated 16-20 hours in a shaking incubator at 37 degrees Celsius. A bacterial stock of these clones was prepared by transferring 100 microliters of the 1 ml cultures to another 96-well plate containing 200 microliters of LB media supplemented with ampicillin. The remaining cells were pelleted by centrifugation and the pellets resuspended in 200 microliters of LB media. One hundred microliters of the concentrated cells were transferred to a 96-well thermowell PCR plate (Costar), and the cells were once again pelleted. The pelleted cells were resuspended in lysis buffer [250 mM Tris-HCl, pH 8.0, 50 mM EDTA, pH 8.0, 8% sucrose, 5% Triton X-100, 1 mM tartrazine, and 666 micrograms per milliliter lysozyme], and the plates were covered with thermowell lids (Costar) and incubated in a MJ Research thermocycler for 2 minutes at 100 degrees Celsius followed by 2 minutes at 25 degrees Celsius. Cell debris was pelleted by centrifugation, and 15 microliters of the supernatant containing the plasmid DNA was electrophoresed on a 0.6x TBE 0.8% agarose gel with appropriate supercoiled size standards to estimate the size of each clone.

The bacterial stocks of clones with inserts in the 5-8 kb size range were used to inoculate 3 ml cultures of LB media supplemented with ampicillin, which were incubated overnight in a shaking incubator at 37 degrees Celsius. Plasmid DNA was prepared from these cultures using the Autogen robotic plasmid preparation machine (Integrated Separation Systems). The resulting DNA templates are subjected to DNA sequencing from both ends with M13 forward or reverse fluorescent energy transfer (FET) dye-labeled primers on ABI 377 sequencers.

DNA sequencing gel files were examined for lane tracking accuracy and adjusted where necessary before data extraction. ABI sample files resulting from gel files were converted to the Standard Chromatogram Format (SCF) [Dear and Staden] and trimmed of sequencing vector (pMYG1 or pMYG2). Trimmed sequences were assembled using Acem.bly (Thierry-Mieg *et al.*, 1995; Durbin and Thierry-Mieg, 1991). Contiguous sequence resulting from automatic assembly was screened for residual vector sequence (both sequencing vector and cloning vector) as well as for bacterial contamination using BLAST (Altschul *et al.*, 1990).

Remaining sequences were arranged according to the relative position and orientation of assembled Sau3AI partial digest clone sequence reads as well as sequence similarity to overlapping genomic clones. Repetitive sequence was masked from the sequence contigs using xblast (Claverie and States, 1993). These masked sequences were placed in a Genetic Data Environment (GDE) (Smith *et al.*, 1994) local database for subsequent similarity searches. Similarities among genomic DNA sequences and hybrid-selected cDNA clones as well as GenBank entries--both DNA and protein--were identified using BLAST. DNA sequences were also characterized with respect to short period repeats, CpG content, and long open reading frames.

EXAMPLE 5

Hybrid selection

Two distinct methods of hybrid selection were used in this work.

Method 1: cDNA preparation and selection. Poly (A) enriched RNA from human mammary gland, prostate, testis, fetal brain, and placenta tissues and from total RNA of the cell line Caco-2 (ATCC HTB 37) were reverse transcribed using the tailed random primer $RXGN_6$ and M-MLV Reverse Transcriptase (Life Technologies, Inc.). First strand cDNA was poly(A) tailed, 2nd strand synthesis was primed with the oligo $RXGT_{12}$, and then the ds cDNA was expanded by amplification with the primer RXG. Hybrid selection was carried out for two consecutive rounds of hybridization to immobilized BAC, PAC or gel purified YAC DNA as described previously. [Parimoo *et al.*, 1991; Rommens *et al.*, 1994]. Individual gel purified YACs or groups of two to four overlapping BAC and/or PAC clones were used in individual selection experiments. Hybridizing cDNA was collected, passed over a G50 Fine Sephadex column and amplified using tailed primers. The products were then digested with EcoRI, size

selected on agarose gels, and ligated into pBluescript (Stratagene) that had been digested with EcoRI and treated with calf alkaline phosphatase (Boehringer Mannheim). Ligation products were transformed into competent DH5α *E. coli* cells (Life Technologies, Inc.).

Characterization of Retrieved cDNAs. 200 to 300 individual colonies from each ligation (from each 250 kbases of genomic DNA) were picked and gridded into microtiter plates for ordering and storage. Cultures were replica transferred onto Hybond N membranes (Amersham) supported by LB agar with ampicillin. Colonies were allowed to propagate and were subsequently lysed with standard procedures. Initial analysis of the cDNA clones involved a prescreen for ribosomal sequences and subsequent cross screenings for detection of overlap and redundancy.

Approximately 10-25% of the clones were eliminated as they hybridized strongly with radiolabeled cDNA obtained from total RNA. Plasmids from 25 to 50 clones from each selection experiment that did not hybridize in prescreening were isolated for further analysis. The retrieved cDNA fragments were verified to originate from individual starting genomic clones by hybridization to restriction digests of DNAs of the starting clones, of a hamster hybrid cell line that contains chromosome 1 as its only human material, and to human genomic DNA. The clones were tentatively assigned into groups based on the overlapping or non-overlapping intervals of the genomic clones.

Method 2: cDNA Preparation. Poly(A) enriched RNA from human mammary gland, fetal brain, lymphocyte, pancreas, prostate, stomach, and thymus were reverse-transcribed using the tailed random primer XN_{12} and Superscript II reverse transcriptase (Gibco BRL). After second strand synthesis and end polishing, the ds cDNA was purified on Sepharose CL-4B columns (Pharmacia). cDNAs were "anchored" by ligation of a double-stranded oligo RP (RP-2 annealed to RL-1) to their 5' ends (5' relative to mRNA) using T4 DNA ligase. Anchored ds cDNA was then repurified on Sepharose CL-4B columns.

Selection was performed by a modified procedure of Lovett et al. (1991). cDNAs from mammary gland, fetal brain, lymphocyte, pancreas, prostate, stomach, and thymus tissues were first expanded by amplification using a nested version of RP, RP.A and XPCR, and purified by fractionation on Sepharose CL-4B. Selection probes were prepared from purified P1s, BACs or PACs by digestion with HinfI and Exonuclease III. The single-stranded probe was photolabelled with photobiotin (Gibco BRL) according to the manufacturer's recommendations. Probe, cDNA and C_0t-1 DNA and poly A DNA were hybridized in 2.4M TEA-Cl, 10mM

$NaPO_4$, 1mM EDTA. Hybridized cDNAs were captured on streptavidin-paramagnetic particles (Dynal), eluted, and reamplified with a further nested version of RP, RP.B and XPCR, and gel purified. The selected, amplified cDNA was hybridized with an additional aliquot of probe, C_0t-1 DNA and poly A DNA. Captured and eluted products were amplified again with RP.B and XPCR, size-selected by gel electrophoresis, and cloned into dephosphorylated HincII cut pUC18. Ligation products were transformed into XL2-Blue ultra-competent cells (Stratagene).

Both methods: Insert-containing clones were identified by blue/white selection on Xgal or Bluo-gal plates. Inserts were amplified by colony PCR with vector primers and then sequenced on ABI 377 sequencers. Alignment of these cDNA sequences to corresponding genomic sequences, and parsing of the revealed exons across those genomic sequences, allowed initial characterization of genes located within the region.

Table 5

Oligonucleotides Used for Hybrid Selection

Name	Sequence	Sequence ID#
$RXGN_6$	5'-CGGAATTCTGCAGATCTA'B'C'NNNNNN	57
$RXGT_{12}$	5'-CGGAATTCTGCAGATCTTTTTTTTTTTT	58
RXG	5'-CGGAATTCTGCAGATCT	59
XN_{12}	5'-(NH2)-GTAGTGCAAGGCTCGAGAACNNNNNNNNNNNN	60
RP-2	5'-(NH2)-TGAGTAGAATTCTAACGGCCGTCATTGTTC	61
RL-1	5'-GAACAATGACGGCCGTTAGAATTCTACTCA-(NH2)	62
RP.A	5'-TGAGTAGAATTCTAACGGCCGTCAT	63
XPCR	5'-(PO4)-GTAGTGCAAGGCTCGAGAAC	64
RP.B	5'-(PO4)-TGAGTAGAATTCTAACGGCCGTCATTG	65

EXAMPLE 6

Inter-exon PCR and RACE for the identification of new exons (5', 3', or internal) of the HPC1 gene

Inter-exon PCR: Following sequence analysis of the first three hybrid selected clones that originated from HPC1, several primers were designed to try to amplify HPC1 products from fetal brain, breast, pancreas, prostate, stomach, and thymus cDNAs. Two important pieces of data were revealed by this experiment: (1) The transcript is not abundant, but it was considerably more abundant in prostate and thymus cDNA than in the other tissues tested. (2) The transcript is subject to a complex pattern of alternative splicing. Specifically, amplification from 1 ng of

prostate and 1 ng of thymus cDNA with the primers 07CG01#F1 and 07CG01#BR1 (see Table 6) and TaqPlus DNA polymerase (Stratagene) revealed more than 8 distinct splice variants of the HPC1 transcript. Amplification was by hot start PCR; conditions used were an initial denaturation step at 95°C for 30 sec followed by a pause at 80°C while the polymerase/ nucleotide mixture was added to the template/primer mixtures. The hot start was followed by 35 cycles of denaturation at 96°C (4 s), annealing at 60°C (10 s) and extension at 72°C (60 s). The bands representing these splice variants were plugged, reamplified, and sequenced using dye terminator chemistry on ABI 377 sequencers. Parsing of these cDNA sequences across the genomic sequence of HPC1 revealed several new exons.

5' RACE: The 5' end exons of HPC1 were identified by a modified RACE protocol called biotin capture 5' RACE (Tavtigian et al., 1996). Poly(A) enriched RNA from prostate was reverse-transcribed using the tailed random primer XN_{12} and Superscript II reverse transcriptase (Gibco BRL). After second strand synthesis and end polishing, the ds cDNA was purified on Sepharose CL-4B columns (Pharmacia). cDNAs were "anchored" by ligation of a double-stranded oligo RP (RP-2 annealed to RL-1) to their 5' ends (5' relative to mRNA) using T4 DNA ligase. Anchored ds cDNA was then repurified on Sepharose CL-4B columns.

The 5' sequences of HPC1 were amplified using two primer combinations: (1) biotinylated reverse primer 07CG01#BR1 (See Table 6) and RP.A, and (2) biotinylated reverse primer 07CG01#BR2 (See Table 6) and RP.A. PCR products were fractionated on an agarose gel, gel purified, and captured on streptavidin-paramagnetic particles (Dynal). Material captured after amplification with 07CG01#BR1 and RP.A was reamplified using the nested phosphorylated reverse primer 07CG01#PR2 and a further nested version of RP-2, RP.B. Material captured after amplification with 07CG01#BR2 and RP.A was reamplified using the nested phosphorylated reverse primer 07CG01#PR3 and RP.B. These PCR reactions gave several bands on an agarose gel; the PCR products were gel purified and sequenced in the reverse direction, using primer 07CG01#PR2 and/or 07CG01#PR3 with dye terminator chemistry on an ABI 377 sequencer.

3' RACE: A 3' end exon of HPC1 was identified by a modified RACE protocol called biotin capture 3' RACE. Poly(A) enriched RNA from prostate was reverse-transcribed using the tailed random primer XT_{15} and Superscript II reverse transcriptase (Life Technologies). The first strand (heteroduplex) cDNA was purified by fractionation on a Sepharose CL-6B column.

The 3' sequence of HPC1 was amplified with the biotinylated forward primer 07CG01#BF4 and the anchor primer XPCR. PCR products amplified with these primers were fractionated on an agarose gel, gel purified, and captured on streptavidin-paramagnetic particles (Dynal). Captured material was reamplified using the nested phosphorylated forward primer 07CG01#PF5 and XT4.

PCR products were gel purified, ligated into the vector pMYG2, and transformed into DH5a cells. Colony PCR products were sequenced using primer 07CG01#PF5 and XPCR using dye terminator chemistry on an ABI 377 sequencer.

Table 6

Oligonucleotides Used for RACE

Name	Sequence	Sequence ID#
XT_{15}	5'-(NH_2)-GTAGTGCAAGGCTCGAGAACTTTTTTTTTTTTTT	66
XT_4	5'-(PO4)-GTAGTGCAAGGCTCGAGAACTTTT	67
XN_{12}	5'-(NH2)-GTAGTGCAAGGCTCGAGAACNNNNNNNNNNNN	68
RP-2	5'-(NH2)-TGAGTAGAATTCTAACGGCCGTCATTGTTC	69
RL-1	5'-GAACAATGACGGCCGTTAGAATTCTACTCA-(NH2)	70
RP.A	5'-TGAGTAGAATTCTAACGGCCGTCAT	71
XPCR	5'-(PO4)-GTAGTGCAAGGCTCGAGAAC	72
RP.B	5'-(PO4)-TGAGTAGAATTCTAACGGCCGTCATTG	73
07CG01#F1	5'-AGG AAG TAT ATC TAA GTC ACC TCC A	74
07CG01#BR1	5'-(Biotin)-AA TTC CAG ACA GAT TGC AGG CAC	75
07CG01#PR2	5'-(PO4)-AG AGG ACT TGT TCC CCA TAA TTG	76
07CG01#BR2	5'-(Biotin)-AG AGG ACT TGT TCC CCA TAA TTG	77
07CG01#PR3	5'-(PO4)-TT ACG GCT ACT GGA GGT GAC TTA	78
07CG01#PR4	5'-(PO4)-AA GTC TCC AGG GCA CAT CTG A	79
07CG01#BF4	5'-(Biotin)-GAAGAAAGAACACTCAGATGTGC	80
07CG01#PF5	5'-(P_{04})-CGAAGGAAAGCTTCCAATTATG	81

EXAMPLE 7

cDNA library screening

Radioactive probes prepared from two hybrid selected clones representative of HPC1 transcripts (mH179o12-4B03 and mH179o12-3B06) were used as probes to screen a total of 5.5 x 10^6 recombinant phage from a human prostate λgt11 cDNA library (HL1131b, Clontech). Prehybridization and hybridization was performed at 42°C in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhardt's mixture, 0.2 mg/ml denatured salmon sperm DNA and 2 mg/ml poly (A).

Dextran sulfate (4% v/v) was included in the hybridization solution only. The filters were rinsed in 2X SSC for 10 minutes at room temperature and then rinsed in 2X SSC/0.1% SDS for 30 minutes at 60°C followed by two washes in 1X SSC/0.1% SDS for 20 minutes each at 60°C. The positive phage were retested for second and third screenings, as required, to obtain purified plaques for sequencing. Inserts were amplified by phage PCR with vector primers and then sequenced using dye terminator chemistry on ABI 377 sequencers.

EXAMPLE 8

Mutation screening

Both genomic DNA and cDNA were used as templates for mutation screening.

Genomic DNA: Using genomic DNAs from prostate kindred members, prostate cancer affecteds, and tumor cell lines as templates, nested PCR amplifications were performed to generate PCR products of the candidate genes that were screened for HPC1 mutations. The primers listed in Table 7 were used to produce amplicons of the HPC1 gene. Using the outer primer pair for each exon (FA-RP, i.e., forward A and reverse P), 1-10 ng of genomic DNA were subjected to a 23-26 cycle primary amplification, after which the PCR products were diluted 60-fold and reamplified using nested M13-tailed primers (FB-RQ, FC-RR or FB-RR) for another 20-25 cycles; either TaqPlus (Stratagene) or AmpliTaq Gold (Perkin Elmer) was used in the PCRs. In general, the PCR conditions used were an initial denaturation step at 95°C for 1 min (TaqPlus) or 10 min (AmpliTaq Gold), followed by cycles of denaturation at 96°C (12 s), annealing at 55°C (15 s) and extension at 72°C (45-60 s). PCR products were sequenced with M13 forward or reverse fluorescent energy transfer (FET) dye-labeled primers on ABI 377 sequencers. Chromatograms were analyzed for the presence of polymorphisms or sequence aberrations in either the Macintosh program Sequencer (Gene Codes) or the Java program Mutscreen (Myriad, proprietary).

Table 7

Oligonucleotides Used for Mutation Screening from Genomic DNA

Mut Amplicon	Name	Sequence	SEQ ID NO:
CA7.CG1. m1	ca7.CG01. m1 A	GTA ATG AAA TCT GAG AAG CTG AA	82
	ca7.CG01. m1 A.a	CAC ACA GTG GTT AAT CAT AAA TAC	83
	ca7.CG01. m1P	CAC AAA GGT ATC TTT TAA GTT CC	84
	ca7.CG01. m1B	GTT TTC CCA GTC ACG ACG GAA GCT GAA TTT AGC AAT ACA GA	85
	ca7.CG01. m1B.a	GTT TTC CCA GTC ACG ACG TTA TCT GTT CAC TTC ACC TTT G	86
	ca7.CG01. m1Q	AGG AAA CAG CTA TGA CCA TCC TGA GCT TTC AAA AAA GTA TTC	87
	ca7.CG01. m1Q.a	AGG AAA CAG CTA TGA CCA TGG TCT TCA CTT TTC ATT TAC TTC	88
CA7.CG1. m2	CA7.CG01.m2 A	TAG CAT TGT TTG AAG CCA CAG	89
	CA7.CG01.m2 P	CTG GAA GAA ACC TGT AAC TTG	90
	CA7.CG01.m2 B	GTT TTC CCA GTC ACG ACG TGA AGC CAC AGA GTT TTA GAG	91
	CA7.CG01.m2 Q	AGG AAA CAG CTA TGA CCA TTG TTC TCA AAT AAT GTC CCA AA	92
CA7.CG1. m3	CA7.CG01.m3 A	GTA ATG CTA TAA TGT TTG AAA GG	93
	CA7.CG01.m3 P	TTC AGG CTA ACT TCC ATC TTC	94
	CA7.CG01.m3 B	GTT TTC CCA GTC ACG ACG GGT TAC CCC AAC ATA CCT ATG	95
	CA7.CG01.m3 Q	AGG AAA CAG CTA TGA CCA TAA ATA GCA TAC ATA ATG TTT ATT C	96
CA7.CG1. m4	ca7.CG01. 4A	CAA AGA GTA TGG GAG GCT GA	97
	ca7.CG01. 4P	ACT TCA GAG AAC AAC TTC GTC C	98
	ca7.CG01. 4B	GTT TTC CCA GTC ACG ACG GGC TGA GAC TGA CTT GAC TAT T	99
	ca7.CG01. 4Q	AGG AAA CAG CTA TGA CCA TGA GGG TCC ATG AGG CTT C	100
CA7.CG1. m5	ca7cg1.m5 A	GTG AAT GGC TAG ATC CCC TTT	101
	ca7cg1.m5 P	AAT GAA CCT ACA GTG AGG CAG	102
	ca7cg1.m5 B	GTT TTC CCA GTC ACG ACG AAA GAC AAC CAC TCT AAT GTG C	103

	ca7cg1.m5 Q	AGG AAA CAG CTA TGA CCA TGT TCT TTT ACA TCT TAA CCC AG	104
CA7.CG1.m6	ca7cg1.m6 A	TCT AGT CAG CCT TCT TGA AC	105
	ca7cg1.m6 P	GAC GTA ACA GCT AAA ACG AA	106
	ca7cg1.m6 B	GTT TTC CCA GTC ACG ACG CCT TCT TGA ACT AGA ACT TG	107
	ca7cg1.m6 Q	AGG AAA CAG CTA TGA CCA TCA GGG TTT ATC CTT ATG AA	108
	ca7cg1.m6 C	GTT TTC CCA GTC ACG ACG TCA CAT GCT CAA AAT CTA AA	109
	ca7cg1.m6 R	AGG AAA CAG CTA TGA CCA TAA GGC AAT CTT TCC AGT G	110
CA7.CG1.m7	ca7cg1.m7 A	CTG AAT TGG GGT TTG TCT TG	111
	ca7cg1.m7 P	AAA GAA AGC AGA ACC TTA GC	112
	ca7cg1.m7 B	GTT TTC CCA GTC ACG ACG TTC TCC TTA CCA TTA GAG CA	113
	ca7cg1.m7 Q	AGG AAA CAG CTA TGA CCA TAT AGG TGG CCT TGT TAT GTA	114
PCR+seq.	m7f2.ca7cg1	TTC TCC TTA CCA TTA GAG CAC	115
PCR	FAM-m7r1.ca7cg1	[FAM]-CC TTC GGA TTT GTT CAA GTC	116
CA7.CG1.m8	ca7cg1.m8 A	CCA TTT GCC TAA TGA ATG AA	117
	ca7cg1.m8P	GTC AGA AAA TCT TGG GTG TA	118
	ca7cg1.m8 B	GTT TTC CCA GTC ACG ACG CTT AAG AAA GAG ATT GCC A	119
	ca7cg1.m8 Q	AGG AAA CAG CTA TGA CCA TGC AAT GTG GTA TTA CAA CTT A	120
	ca7cg1.m8 C	GTT TTC CCA GTC ACG ACG AAA ATA AGC TGT CTC TGA AG	121
	ca7cg1.m8 R	AGG AAA CAG CTA TGA CCA TGG GTG TAA AAT AAT TTC TGG	122
CA7.CG1.m9	ca7cg1.A	CGT CTT ACT CAG TTT TGT ATT CT	123
	ca7cg1.P	CAT CTA GAA GTA TGC ATT TGG TA	124
	ca7cg1.B	GTT TTC CCA GTC ACG ACG TGA ATC TTA TTT TCT GCA AGG C	125
	ca7cg1.Q	AGG AAA CAG CTA TGA CCA TTC AAA TAA GGT ATA AAG ACA GAG	126
	ca7cg1.C	GTT TTC CCA GTC ACG ACG AAT CCC TGA ATG GAT AGC ACC C	127
	ca7cg1.R	AGG AAA CAG CTA TGA CCA TAA ATC	128

		ACA AAA ATG TCT AAG GTT	
CA7.CG1.m10	ca7cg1.m10A	CTG AAT CTC CCC TAT TAG AAG T	129
	ca7cg1.m10P	AAG GCC ATT AA GAG GTT CTT AG	130
	ca7cg1.m10B	GTT TTC CCA GTC ACG ACG GAG TTA CAT TCA TTT TTC GAG TC	131
	ca7cg1.m10Q	AGG AAA CAG CTA TGA CCA TTT CAA GAC CAG CCT GAC CAA C	132
CA7.CG1.m11	ca7cg1.m11A	TCC CTG TTG AAA TTC CAA CCT	133
	ca7cg1.m11P	CAT AGA AAT TCT CAC CTA CCC A	134
	ca7cg1.m11B	GTT TTC CCA GTC ACG ACG CCA AGG TGA TGG TAT GTA GAG	135
	ca7cg1.m11Q	AGG AAA CAG CTA TGA CCA TTG TAA ATG GAT CTT GAA GAT CAT	136
CA7.CG1.m12	ca7cg1.m12A	GCA CAG AGC ACA TTC TGG TGA	137
	ca7cg1.m12P	TCC CAA AGA AAA CTA CTA GCC	138
	ca7cg1.m12B	GTT TTC CCA GTC ACG ACG CTG ATG ATC ACA GTC TCT AAG	139
	ca7cg1.m12Q	AGG AAA CAG CTA TGA CCA TCC AGC AAA GTT GTT GTT GGTT	140
CA7.CG1.m13	ca7.CG01.13A	AGA CAG TTG GTA TTT AGG GA	141
	ca7.CG01.13P	TCA TTA TTG CAT TTT CTG GA	142
	ca7.CG01.13B	GTT TTC CCA GTC ACG ACG AGC CAT TTT CCT CTC TCC A	143
	ca7.CG01.13Q	AGG AAA CAG CTA TGA CCA TGG GCT TCT TTT CCA CTT CAA	144
CA7.CG1.m14	ca7cg1.m14A	CAA CCA AAC TAT TAT GAA ACC G	145
	ca7cg1.m14P	AGT GGG GAG CCA GTG CTG TTA	146
	ca7cg1.m14B	GTT TTC CCA GTC ACG ACG TTA TAA TAA TCA CTA GAG ATA GG	147
	ca7cg1.m14Q	AGG AAA CAG CTA TGA CCA TAA TCT TGT ATG TTC TCC CAG G	148
CA7.CG1.m15	ca7cg1.m15A	TTG GTG GCA GTA GAC TGT GGT	149
	ca7cg1.m15P	GAC AGC TAT TAC TCA AAT GTC A	150
	ca7cg1.m15B	GTT TTC CCA GTC ACG ACG TAA GAT TTT GCT ACG CAA ACT GT	151

	ca7cg1.m15Q	AGG AAA CAG CTA TGA CCA TAG AGA CCC GAG TAA GCA TAG T	152
CA7.CG1.m16	ca7cg1.m16A	TGG ACA AGT CAA TGC ACT ACT G	153
	ca7cg1.m16P	TGA TTT AAG CTG CCC AGA TTT C	154
	ca7cg1.m16B	GTT TTC CCA GTC ACG ACG TCT TCT TTA GTT GAG AGA ACC T	155
	ca7cg1.m16Q	AGG AAA CAG CTA TGA CCA TGG AGC CAT GTT GGG CAC AGT	156
	ca7cg1.m16C	GTT TTC CCA GTC ACG ACG ACA GCT ATG AAA TAG AAC AGA G	157
	ca7cg1.m16R	AGG AAA CAG CTA TGA CCA TGC ATA CGT GCA GCA ACA GAG A	158
CA7.CG1.m17	ca7cg1.m17a1	TTG GTC TCA GAA ATA ATC TTA CTG G	159
	ca7cg1.m17p1	GGA TGT AGC ACC TTG AAA TCA TTC	160
	ca7cg1.m17b1	GTT TTC CCA GTC ACG ACG AGC CTA TGG ATG TAT TTA TTC AGT TA	161
	ca7cg1.m17q1	AGG AAA CAG CTA TGA CCA TGT TCC ATT CGT TTC CTA TCA TTA G	162
ca7cg1.m18	ca7cg1.m18A	GGC AAA AAA ATC AAT AAT ATG	163
	ca7cg1.m18P	CAT TGC CCA CCT GTC TAA C	164
	ca7cg1.m18B	GTT TTC CCA GTC ACG ACG AAG ATT GTT AAA TGC TAC TGC	165
	ca7cg1.m18Q	AGG AAA CAG CTA TGA CCA TTA TCA CTA TTC CCC TTG GC	166
ca7cg1.m19	ca7cg1.m19A	GGA ATG TGG AGT AAT GTA AAC	167
	ca7cg1.m19P	CAC CAT GTT GAA ATT AAG CAG	168
	ca7cg1.m19B	GTT TTC CCA GTC ACG ACG GTA ATT GTT GAT AGT CCT CTG	169
	ca7cg1.m19Q	AGG AAA CAG CTA TGA CCA TCA TAA AAC CAA AGC ATC CG	170
ca7cg1.m20	ca7cg1.m20A	ATT TGC TGT CAC ATT ACC CTG	171
	ca7cg1.m20P	CAG CCT GCC TGG GTG ACA G	172
	ca7cg1.m20B	GTT TTC CCA GTC ACG ACG TGT CAC ATT ACC CTG TTT ATC	173
	ca7cg1.m20Q	AGG AAA CAG CTA TGA CCA TTA AGA AGA GGT GAT ATT ACT TAC	174

ca7cg1.m21	ca7cg1.m21A	CTA TTG TAA TGA ATG CTG CTG	175
	ca7cg1.m21P	CAG AAG ATT ATC GTG GTC ATC	176
	ca7cg1.m21B	GTT TTC CCA GTC ACG ACG ATC AAG TGA CTC CTA ACC CTG	177
	ca7cg1.m21Q	AGG AAA CAG CTA TGA CCA TCG TGG TCA TCA TAA ACT AAA TAC	178
ca7cg1.m22	ca7cg1.m22A	AAC TTT GAG TCT GTA GGT TGT TC	179
	ca7cg1.m22P	AGA TGA GCA GCC CAC TAT TG	180
	ca7cg1.m22B	GTT TTC CCA GTC ACG ACG CCA TTT GTT GAA GAA AAG TTA AG	181
	ca7cg1.m22Q	AGG AAA CAG CTA TGA CCA TCA GAA AAG GCT GGA CAA CTT G	182
	ca7cg1.m22 C1	GTT TTC CCA GTC ACG ACG CAA CTA TTC ATC TCT TAT CTA CC	183
	ca7cg1.m22 R1	AGG AAA CAG CTA TGA CCA TTG AGC AGC CCA CTA TTG ATT TC	184
ca7cg1.m23	ca7cg1.m23A	GAA TGG AAT AAG TTA AAT CTT TG	185
	ca7cg1.m23P	TAT CTG AAA AAC TAA TAA GCC AG	186
	ca7cg1.m23B	GTT TTC CCA GTC ACG ACG TTG CTT TCT ACT CAG AGT CTA TG	187
	ca7cg1.m23Q	AGG AAA CAG CTA TGA CCA TAC TAA CAT AAT TGG CTA ATG GC	188
ca7cg1.m24	ca7cg1.m24A	CAG GAT TAT ACT TTC ACT CAA G	189
	ca7cg1.m24P	GAC ATT TAA CTT AAT TTC ACT TG	190
	ca7cg1.m24B	GTT TTC CCA GTC ACG ACG ATA GAC TCA AGA AAA ATG CTA AG	191
	ca7cg1.m24Q	AGG AAA CAG CTA TGA CCA TCT CCT TGT TAT TTC TAA ACC AG	192
ca7cg1.m25	ca7cg1.m25A	TTG TCT ACC TGA ACC CCG AG	193
	ca7cg1.m25P	CAA AAT GGG GCT TGA TTA GG	194
	ca7cg1.m25B	GTT TTC CCA GTC ACG ACG TAC CTT TCT GTG CGT GAT AGC	195
	ca7cg1.m25Q	AGG AAA CAG CTA TGA CCA TTT AGG GCT CAA ACT GAA ATG G	196

cDNA: Total RNA prepared from either tumor cell lines or prostate kindred lymphocytes was treated with DNase I (Boehringer Mannheim) to remove contaminating genomic DNA, and then reverse transcribed to heteroduplex cDNA with a mix of N_{10} random primers and a tailed oligo dT primer, and Superscript II reverse transcriptase (Life Technologies). This cDNA was used as the template for nested PCR amplifications to generate the cDNA PCR products of the candidate genes that were screened for HPC1 mutations. Using the outer primer pair for each amplicon, 10 ng of cDNA were subjected to a 20 cycle primary amplification, after which the PCR products were diluted 100-fold and reamplified using nested M13-tailed primers for another 25-30 cycles. The cDNAs were amplified by hot start PCRs using TaqPlus DNA polymerase (Stratagene). Conditions used were an initial denaturation step at 95°C for 30 sec followed by a pause at 80°C while the polymerase/nucleotide mixture was added to the template/primer mixtures. The hot start was followed by cycles of denaturation at 96°C (4 s), annealing at 55°C (10 s) and extension at 72°C (60 s). PCR products were gel purified and then sequenced with M13 forward or reverse fluorescent energy transfer (FET) dye-labeled primers on ABI 377 sequencers. The sequences of these products were analyzed in GDE to determine their exon structure. Chromatograms were analyzed for the presence of polymorphisms or sequence aberrations in either the Macintosh program Sequencher (Gene Codes) or the Java program Mutscreen (Myriad, proprietary).

Table 8

Oligonucleotides Used for Mutation Screening from cDNA

Name	Sequence	Sequence ID#
ca7.CG01.13C	AGCCATTTTCCTCTCTCCA	197
ca7.CG01.13D	GTTTTCCCAGTCACGACGCCACCACATACCACACTTC	198
ca7.CG01.1C	CAGAATCGCATCAGTAATAGA	199
ca7.CG01.1D	GTTTTCCCAGTCACGACGTGAAGACCTCTTTGAATTATC	200
ca7.CG01.2R	GAAGCTGTGTTCTTTTTTCA	201
ca7.CG01.2S	AGGAAACAGCTATGACCATCTGTGTTCTTTTTTCAGTAGTTA	202

EXAMPLE 9

Analysis of HPC1 Mutations

The DNA samples which were screened for HPC1 mutations were extracted from blood or tumor samples from patients with prostate or ovarian cancer (or known carriers by haplotype analysis) who were participating in research studies on the genetics of prostate cancer. All subjects signed appropriate informed consent.

Role of HPC1 in Cancer. Most tumor predisposition genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense; some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wild-type product (Shaulian *et al.* 1992; Srivastava *et al.*, 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su *et al.*, 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little *et al.*, 1993).

Sequence for HPC1 has been determined. Twenty five exons have been sequenced and several of the alternative splice variants of the transcript have been determined. SEQ ID NOs:1-52 show the sequence for HPC1 including exons and flanking genomic sequence. The exon names in their order in genomic DNA sequence in the direction of transcription of the gene are as follows: g1m20, g2m13, g3m1, [g4m17a, g4m17b], g5m19, g6m18, g7m10, g8m11, g9m21, g10m2, g11m3, g12m14, g13m22, g14m12, g15m4, g16m23, g17m24, g18m25, [g19m5a, g19m5b] g20m6, g21m7, g22m15, g23m9, g24m16, g25m8 The remaining sequences include the same exons plus surrounding intron, e.g., SEQ ID NO:2 includes SEQ ID NO:1 within it. SEQ ID NOs:7 and 8 are alternate forms of a single exon and both are included within SEQ ID NO:9. Similarly, SEQ ID NOs:38 and 39 are alternate forms of a single exon and both are included within SEQ ID NO:40.

Certain rules have been determined concerning splicing. The transcripts must begin with the exon g1m20 or g2m13. Only 1 of these two exons is present, e.g., if the exon represented by glm20 is present, then the exon represented by g2m13 is not present. The transcript must terminate with one of the exons represented by g4m17a, g22m15, g24m16 or g25m8. Exons represented by g4m17a and g4m17b] are alternate forms and only one of these two forms of the exon may be present, e.g., if the exon represented by g4m17a is present then the exon represented by g4m17b is absent. Similarly, exons represented by g19m5a, g19m5b are

alternate forms and only one of these two forms of the exon may be present, e.g., if the exon represented by g19m5a is present then the exon represented by g19m5b is absent.. The exons represented by g4m17a, g22m15, g24m16 have poly A sequences.

In studying the several kindreds, two mutations were found which were associated with cancer. These are shown in Table 9.

Also found were two polymorphisms which are in disequilibrium as shown in Table 10. These occur at base 207 of SEQ ID NO:2 and at base 158 of SEQ ID NO:10.

There are many potential combinations of exons which could be spliced together to form an HPC1 transcript. Following the rules described above (as determined from sequencing many transcripts), examples of the combinations of spliced exons are shown in Table 11. These examples are not inclusive of all combinations of exons which have been found and which are possible. Many additional combinations exons are determined by applying the above rules.

SEQ ID NOs:203-210 show some putative polypeptides which are obtained from some of the mRNA variants.

Table 9

Mutations in HPC1

Sample	Type	Exon Name	SEQ ID NO:	Sequence Alteration	Reading Frame	Coding Alteration
Tumor cell line SW 48 (CCL-231)	TCL	g3m1	5	del.A@x61/het		frameshift
Tumor cell line DBTRG-05MG (CRL-2020)	TCL	g19m5a,b	38,39	C>T@x31	1	His>Tyr
Tumor cell line LNCAP.FGC (CRL-1740)	TCL	g21m7	43	del.T@x9/het		frameshift
Tumor cell line SCaBer (HTB-3)	TCL	g23m9	47	del.T@x81/het		frameshift
Tumor cell line WiDr (CCL-218)	TCL	g23m9	47	C+T@x151	1 or 2 ?	Gln>Stop or Cys>Cys
Tumor cell line HT-29 (HTB-38)	TCL	g23m9	47	C>A@x151	1 or 2 ?	Gln>Lys or Cys>Stop
Prostate tumor/normal pair Canji 14 T/N	Prostate Tumor/Normal Pair	g23m9	47	G>A@x42	1 or 2 ?	Trp>Ter or Gly>Glu
Tumor cell line NCI-H676B (HTB-179)	TCL	g8m11	16	C+T@x21	1 or 2 ?	Ser>Phe or Pro>Ser
Tumor cell line ZR-75-30 (CRL-1504)	TCL	g12m14	24	C+T@x22	1 or 2 ?	Ala>Val or Leu>Phe
Tumor cell line U-87 MG (HTB-14)	TCL	g13m22	26	A>G@x182	1	Gln>Arg

Table 10

Exon	g2m13		g3m1			g10m2		g19m5	g23m9	
Sequence ID #	3	4	5	6	20	19	20	38,39	47	47
Commonly polymorphic position	X64	e/i+5	X68	e/i+38	i/e-7	X66	x/i+20	X31	X42	X73
alleles (more / less common)	G /A	T / A	G / A	C / T	C / G	G / T	A / T	C / T	G / A	C / T
Prostate Tumor Normal Pair										
CanJl 14 T/N	nd	nd	nd	nd	C	G	A / T	nd	G / A (gly->glu)	C
Tumor Cell Lines #1										
BT-549 (HTB-122)	A	T	G	C	C	G	A	C	G	C
MCF7 (HTB-22)	A	T	G	C	C	G	A	C	G	C
MDA-MB-415 (HTB-128)	G	T / A	G	/ C / T	C	G	A	C	G	C
ZR-75-1 (CRL-1500)	G	T	G	T	C	G	A	C	G	C
DU145 (HTB-81)	G	T	G	C	C	G	T	C	G	C
LNCAP.FGC (CRL-1740)	G	A	G	C	C	G	T	C	G	C / T
NCI-H660 (CRL-5813)	G	T	G / A	C	G	G	A	C	G	C
PC-3 (CRL-1435)	G	A	G	C	C	G	T	C	G	C
U-373 MG (HTB-17)	A	T	G	C	C	G	A	C	G	C
U-138 MG (HTB-16)	A	T	G	C	C	G	A	C	G	C
U-87 MG (HTB-14)	G	A	G	C	C	G	T	C	G	C
CCF-STTG1 (CRL-1718)	G	T	G / A	C	C	G	A	C	G	C
T98G (CRL-1690)	G	T / A	G	C	C	G	A / T	C	G	C / T
SW 1783 (HTB-13)	G	T	G	T	C	G	A	C	G	C
DBTRG-05MG (CRL-2020)	G+A	T / A	G	C	C	G	A / T	C / T (his->tyr)	G	C

Table 10 (Con't)

Tumor Cell Lines #2										
HS 578T (HTB-126)	G	A	G	C	C	G	A / T	C	G	C / T
MDA-MB-435S (HTB-129)	G	T	G	C	C	T	A	C	G	nd
T-47D (HTB-133)	G	A	G	C	C	G	A / T	C	G	C
Bx PC-3 (CRL-1687)	G+A	T	G	C	C	G	A	C	G	C / T
CAPAN-1 (HTB-79)	G	T	G	C	C	G	A / T	C	G	nd
CFPAC-1 (CRL-1918)	G	T	G / A	C	C	G	A	C	G	C
COLO 587 (CRL-2000)	G+A	T	G	C / T	C	G	A	C	G	nd
HPAC (CRL-2119)	A	T	G	C	C	G	A	C	G	nd
HPAF-II (CRL-1997)	G	T	G	C / T	C	G	A	C	G	T
Hs 700T (HTB-147)	A	T	G	C	C	G	A	C	G	C
Hs 766T (HTB-134)	G	T	G / A	C	C	G	T	C	G	C
MIA PaCa-2 (CRL-1420)	G	T	G	C	C	G	A	C	G	T
PANC-1 (CRL-1469)	G	T	G / A	C	C	G	A	C	G	C / T
SU.86.86 (CRL-1837)	G	T / A	G / A	C	C	G	A / T	C	G	C / T
CAPAN-2 (HTB-80)	A	T	G	C	C	G	A	C	G	C
ASPC-1 (CRL-1682)	A	T	G	C	C	G	A	C	G	C
allele frequences (120 controls)				.167	.029	.042	.097	.000	nd	nd
allele frequences (>90 controls)	.115	.093	.183							
allele frequences (31 tumor cell	.244	.200	.136	.111	.022	.022	.244	.022	.000	nd
lines, about 45 chromosomes)										
nd = not determined										

Table 11

Splice Variant Example #	Exon content
1.	g1m20 g4m17a
2.	g1m20 g3m1 g4m17a
3.	g1m20 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
4.	g1m20 g3m1 g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
5.	g1m20 g3m1 g4m17b g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
6.	g1m20 g3m1 g4m17b g5m19 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
7.	g1m20 g3m1 g4m17b g5m19 g6m18 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
8.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
9.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
10.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
11.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
12.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8
13.	g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

14. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

15. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

16. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

17. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

18. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g19m5a g20m6 g21m7 g23m9 g25m8

19. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g20m6 g21m7 g23m9 g25m8

20. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g21m7 g23m9 g25m8

21. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g23m9 g25m8

22. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g25m8

23. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

24. g1m20 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

25. g1m20 g3m1 g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

26. g1m20 g3m1 g4m17b g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

27. g1m20 g3m1 g4m17b g5m19 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

28. g1m20 g3m1 g4m17b g5m19 g6m18 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

29. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

30. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

31. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

32. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

33. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

34. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

35. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

36. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

37. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

38. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

39. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g19m5bg20m6 g21m7 g23m9 g25m8

40. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g20m6 g21m7 g23m9 g25m8

41. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g21m7 g23m9 g25m8

42. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g23m9 g25m8

43. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g25m8

44. g1m20 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

45. g2m13 g4m17a

46. g2m13 g3m1 g4m17a

47. g2m13 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

48. g2m13 g3m1 g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

49. g2m13 g3m1 g4m17b g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

50. g2m13 g3m1 g4m17b g5m19 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

51. g2m13 g3m1 g4m17b g5m19 g6m18 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

52. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

53. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

54. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

55. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

56. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

57. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

58. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

59. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

60. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

61. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

62. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g19m5a g20m6 g21m7 g23m9 g25m8

63. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g20m6 g21m7 g23m9 g25m8

64. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g21m7 g23m9 g25m8

65. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g23m9 g25m8

66. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g25m8

67. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5a g20m6 g21m7 g23m9 g25m8

68. g2m13 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

69. g2m13 g3m1 g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

70. g2m13 g3m1 g4m17b g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

71. g2m13 g3m1 g4m17b g5m19 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

72. g2m13 g3m1 g4m17b g5m19 g6m18 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

73. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

74. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

75. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

76. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

77. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

78. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

79. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

80. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

81. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

82. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

83. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g19m5bg20m6 g21m7 g23m9 g25m8

84. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g20m6 g21m7 g23m9 g25m8

85. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g21m7 g23m9 g25m8

86. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g23m9 g25m8

87. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g25m8

88. g2m13 g3m1 g4m17b g5m19 g6m18 g7m10 g8m11 g9m21 g10m2 g11m3 g12m14 g13m22 g14m12 g15m4 g16m23 g17m24 g18m25 g19m5b g20m6 g21m7 g23m9 g25m8

EXAMPLE 10

Polymorphisms in HPC1

In the course of determining the sequence of the HPC1 gene from the many kindred samples as well as tumor and cell line samples, two mutation have been found (see Table 9) as well as several polymorphisms. As noted previously, two of the polymorphisms which have been found to date show linkage disequilibrium. These and other polymorphisms are shown in Table 12. Frequencies have been tested in controls, tumor cell lines (TCLs) and family members of kindreds with cancer.

Table 12

Polymorphisms in HPC1

1) A G/A polymorphism is found located at base 68 of SEQ ID NO:3. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous G - 78	Homozygous G - 26	Homozygous G - 13
Heterozygous - 34	Heterozygous - 6	Heterozygous - 1
Homozygous A - 6	Homozygous A - 0	Homozygous A - 0

2) A C/T polymorphism is found located at base 186 of SEQ ID NO:4. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous C - 85	Homozygous C - 27	Homozygous C - 10
Heterozygous - 29	Heterozygous - 3	Heterozygous - 4
Homozygous T - 4	Homozygous T - 2	Homozygous T - 0

3) A C/G polymorphism is found located at base 29 of SEQ ID NO:10. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members	Prostate – Canji
Homozygous C - 111	Homozygous C - 31	Homozygous C - 14	Not tested
Heterozygous - 7	Heterozygous - 0	Heterozygous - 0	Not tested
Homozygous G - 0	Homozygous G - 1	Homozygous G - 0	Not tested

4) A G/T polymorphism is found located at base 66 of SEQ ID NO:9. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous G - 108	Homozygous G - 31	Homozygous G - 10
Heterozygous - 10	Heterozygous - 0	Heterozygous - 4
Homozygous T - 0	Homozygous T - 1	Homozygous T - 0

5) An A/T polymorphism is found located at base 158 of SEQ ID NO:10. The T allele of this polymorphism occurs at a significantly higher frequency in the TCLs than in the controls and is thus in disequilibrium with "the state of being a tumor cell line." The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous A - 97	Homozygous A - 20	Homozygous A – 10
Heterozygous - 19	Heterozygous - 6	Heterozygous - 4
Homozygous T - 2	Homozygous T - 6	Homozygous T - 0

6) A C/T polymorphism is found located at base 31 of SEQ ID NO:19 and also at base 31 of SEQ ID NO:20 (these are the identical locations). The T variant of this polymorphism is actually a mutation which was seen in a glioma tumor cell line (DBTRG-05MG) and has not been seen elsewhere other than in the homozygous C form. The T allele of this variant results in the non-conservative missense change histidine → tyrosine in the most likely reading frame of this exon.

The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous C - 118	Homozygous C - 31	Homozygous C - 14
Heterozygous - 0	Heterozygous - 1	Heterozygous - 0
Homozygous T - 0	Homozygous T - 0	Homozygous T - 0

7) An A/T polymorphism is found located at base 173 of SEQ ID NO:25. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Not tested	Homozygous A - 24	Homozygous A - 5
Not tested	Heterozygous - 4	Heterozygous - 9
Not tested	Homozygous T - 4	Homozygous T - 0

8) An A/G polymorphism is found located at base 183 of SEQ ID NO:25. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Not tested	Homozygous A - 25	Homozygous A - 14
Not tested	Heterozygous - 3	Heterozygous - 0
Not tested	Homozygous G - 4	Homozygous G - 0

9) A C/T polymorphism is found located at base 73 of SEQ ID NO:28. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Not tested	Homozygous C - 22	Homozygous C - 11
Not tested	Heterozygous - 8	Heterozygous - 3
Not tested	Homozygous T - 2	Homozygous T - 0

10) A G/T polymorphism is found located at base 324 of SEQ ID NO:28. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Not tested	Homozygous G - 30	Homozygous G - 14
Not tested	Heterozygous - 2	Heterozygous - 0
Not tested	Homozygous T - 0	Homozygous T - 0

11) A G/A polymorphism is found located at base 64 of SEQ ID NO:1. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous G - 22	Homozygous G - 21	Homozygous G - 6
Heterozygous - 11	Heterozygous - 3	Heterozygous - 7
Homozygous A - 0	Homozygous A - 8	Homozygous A - 1

12) A T/A polymorphism is found located at base 207 of SEQ ID NO:2. The A allele of this polymorphism occurs at a significantly higher frequency in the TCLs than in the controls and is thus in disequilibrium with "the state of being a tumor cell line." The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Homozygous T - 77	Homozygous T - 22	Homozygous T - 10
Heterozygous - 17	Heterozygous - 4	Heterozygous - 4
Homozygous A - 0	Homozygous A - 5	Homozygous A - 0

13) A G/A polymorphism is found located at base 42 of SEQ ID NO:28. The A variant of this polymorphism is actually a germline mutation which was seen elsewhere other than in the homozygous G form. The A allele of this variant causes the non-conservative missense change glycine → glutamate in the most likely reading frame of this exon. The number of occurrences of the three combinations of polymorphism in the various test groups is:

Controls	TCLs	Family Members
Not tested	Homozygous G - 31	Homozygous G - 14
Not tested	Heterozygous - 0	Heterozygous - 0
Not tested	Homozygous A - 0	Homozygous A - 0

EXAMPLE 11

Analysis of the HPC1 Gene

The structure and function of HPC1 gene are determined according to the following methods.

Biological Studies. Mammalian expression vectors containing HPC1 cDNA are constructed and transfected into appropriate prostate carcinoma cells with lesions in the gene. Wild-type HPC1 cDNA as well as altered HPC1 cDNA are utilized. The altered HPC1 cDNA can be obtained from altered HPC1 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms of the gene.

Molecular Genetics Studies. *In vitro* mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

Mechanism Studies. The ability of HPC1 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

Structural Studies. Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

EXAMPLE 12

Generation of Polyclonal Antibody against HPC1

Segments of HPC1 coding sequence are expressed as fusion protein in *E. coli*. The overexpressed proteins are purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer, *et al.*, 1993).

Briefly, a stretch of HPC1 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, WI). The HPC1 incorporated sequences might include SEQ ID NOs:1, 3, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26, 28, 30 or combinations thereof. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight is verified by SDS/PAGE. Fusion proteins are purified from the gel by electroelution. The identification of the protein as the HPC1 fusion product is verified by protein sequencing at the N-terminus. Next, the purified protein is used as immunogen in rabbits. Rabbits are immunized with 100 mg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 mg of immunogen in incomplete Freund's adjuvant followed by 100 mg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure can be repeated to generate antibodies against mutant forms of the HPC1 protein. These antibodies, in conjunction with antibodies to wild type HPC1, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 13

Generation of Monoclonal Antibodies Specific for HPC1

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact HPC1 or HPC1 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 mg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, MD) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of $2x10^5$ cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of HPC1 specific antibodies by ELISA or RIA using wild type or mutant HPC1

target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 14

Isolation of HPC1 Binding Peptides

Peptides that bind to the HPC1 gene product are isolated from both chemical and phage-displayed random peptide libraries as follows.

Fragments of the HPC1 gene product are expressed as GST and His-tag fusion proteins in both *E. coli* and SF9 cells. The fusion protein is isolated using either a glutathione matrix (for GST fusions proteins) or nickel chelation matrix (for His-tag fusion proteins). This target fusion protein preparation is either screened directly as described below, or eluted with glutathione or imidizole. The target protein is immobilized to either a surface such as polystyrene; or a resin such as agarose; or solid supports using either direct absorption, covalent linkage reagents such as glutaraldehyde, or linkage agents such as biotin-avidin.

Two types of random peptide libraries of varying lengths are generated: synthetic peptide libraries that may contain derivatized residues, for example by phosphorylation or myristylation, and phage-displayed peptide libraries which may be phosphorylated. These libraries are incubated with immobilized HPC1 gene product in a variety of physiological buffers. Next, unbound peptides are removed by repeated washes, and bound peptides recovered by a variety of elution reagents such as low or high pH, strong denaturants, glutathione, or imidizole. Recovered synthetic peptide mixtures are sent to commercial services for peptide micro-sequencing to identify enriched residues. Recovered phage are amplified, rescreened, plaque purified, and then sequenced to determined the identity of the displayed peptides.

Use of HPC1 binding peptides

Peptides identified from the above screens are synthesized in larger quantities as biotin conjugates by commercial services. These peptides are used in both solid and solution phase competition assays with HPC1 and its interacting partners identified in yeast 2-hybrid screens. Versions of these peptides that are fused to membrane-permeable motifs (Lin et al., 1995; Rojas et al., 1996) will be chemically synthesized, added to cultured cells and the effects on growth, apoptosis, differentiation, cofactor response, and internal changes will be assayed.

EXAMPLE 15

Sandwich Assay for HPC1

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle. Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 ml sample (e.g., serum, urine, tissue cytosol) containing the HPC1 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 ml of a second monoclonal antibody (to a different determinant on the HPC1 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., 125-I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of HPC1 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal antibodies which are specific for the wild-type HPC1 as well as monoclonal antibodies specific for each of the mutations identified in HPC1.

EXAMPLE 16

Two-hybrid Assay to Identify Proteins that Interact with HPC1

Sequence encoding all or portions of HPC1 are ligated to pAS2-1 (Clontech) such that the coding sequence of HPC1 is in-frame with coding sequence for the GAL4p DNA-binding domain. This plasmid construct is introduced into the yeast reporter strain Y190 by transformation. A library of activation domain fusion plasmids prepared from human prostate cDNA (Clontech) is then introduced into strain Y190 carrying the pAS2-1-based fusion construct. Transformants are spread onto 20 - 150 mm plates of yeast minimal media lacking leucine, tryptophan, and histidine, and containing 25 mM 3-amino-1,2,4-triazole. After one week incubation at 30°C, yeast colonies are assayed for expression of the lacZ reporter gene by β-galactosidase filter assay. Colonies that both grow in the absence of histidine and are positive for production of β-galactosidase are chosen for further characterization.

The activation domain plasmid is purified from positive colonies by the smash-and-grab technique. These plasmids are introduced into *E. coli* DH5α by electroporation and purified

from *E. coli* by the alkaline lysis method. To test for the specificity of the interaction, specific activation domain plasmids are cotransformed into strain Y190 with plasmids encoding various DNA-binding domain fusion proteins, including fusions to HPC1 and human lamin C. Transformants from these experiments are assayed for expression of the HIS3 and lacZ reporter genes. Positives that express reporter genes with HPC1 constructs and not with lamin C constructs encode bona fide HPC1 interacting proteins. These proteins are identified and characterized by sequence analysis of the insert of the appropriate activation domain plasmid.

This procedure is repeated with mutant forms of the HPC1 gene, to identify proteins that interact with only the mutant protein or to determine whether a mutant form of the HPC1 protein can or cannot interact with a protein known to interact with wild-type HPC1.

LIST OF REFERENCES

Altschul, S.F., *et al.* (1990). *J. Mol. Biol.* **215**: 195-197.

Altschul, S.F., *et al.* (1997). *Nucl. Acids Res.* **25**:3389-3402.

Anand, R. (1992). *Techniques for the Analysis of Complex Genomes*, (Academic Press).

Anderson, *et al.* (1980). *Proc. Natl. Acad. Sci. USA* 77:5399-5403.

Anderson, J.A., *et al.* (1992). *J. Otolaryngology* **21**:321.

Ausubel, F.M., *et al.* (1992). *Current Protocols in Molecular Biology*, (J. Wiley and Sons, NY).

Bandyopadhyay and Temin (1984). *Mol. Cell. Biol.* **4**:749-754.

Bartel, P.L., *et al.* (1993). "Using the 2-hybrid system to detect protein-protein interactions." *In:* Cellular Interactions in Development: A Practical Approach, Oxford University Press, pp. 153-179.

Beaucage and Carruthers (1981). *Tetra. Letts.* **22**:1859-1862.

Berglund, P., *et al.* (1993). *Biotechnology* **11**:916-920.

Berkner (1992). *Curr. Top. Microbiol. Immunol.* **158**:39-61.

Berkner, *et al.* (1988). *BioTechniques* **6**:616-629.

Botstein, *et al.* (1980). *Am. J. Hum. Genet.* **32**:314-331.

Breakfield and Geller (1987). *Mol. Neurobiol.* **1**:337-371.

Brinster, *et al.* (1981). *Cell* **27**:223-231.

Buchschacher and Panganiban (1992). *J. Virol.* **66**:2731-2739.

Cannon, L. et al. (1982). *Cancer Surveys* **1**:47-69.

Capecchi, M.R. (1989). *Science* **244**:1288.

Cariello (1988). *Human Genetics* **42**:726.

Chevray, P.M. and Nathans, D.N. (1992). *Proc. Natl. Acad. Sci. USA* **89**:5789-5793.

Claverie, J.M. and States, D.J. (1993). *Computers and Chemistry* **17**: 191-201.

Conner, B.J., *et al.* (1983). *Proc. Natl. Acad. Sci. USA* **80**:278-282.

Constantini and Lacy (1981). *Nature* **294**:92-94.

Cotten, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:4033-4037.

Cotton, *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:4397-4401.

Cottingham, R..W. et al. (1993). *Am. J. Hum. Genet.* **53**:252-263.

Culver, *et al.* (1992). *Science* **256**:1550-1552.

Curiel, *et al.* (1991a). *Proc. Natl. Acad. Sci. USA* **88**:8850-8854.

Curiel, *et al.* (1991b). *Hum. Gene Ther.* **3**:147-154.

Dear, S. and Staden, R. (1992). *DNA Sequence* 107-110.

Deutscher, M. (1990). *Meth. Enzymology* **182** (Academic Press, San Diego, Cal.).

Donehower, L.A., *et al.* (1992). *Nature* **356**:215.

Durbin, R. and Thierry-Mieg, J. (1991). A *C. elegans* Database. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov.

Enhancers and Eurkaryotic Gene Expression, Cold Spring Harbor Press, Cold Spring Harbor, New York (1983).

Erickson, J. *et al.*, (1990). *Science* **249**:527-533.

Fahy, E., *et al.* (1991). *PCR Methods Appl.* **1**:25-33.

Fain, P.R. (1992). *Cytogen. Cell Genet.* **60**:178.

Felgner, *et al.* (1987). *Proc. Natl. Acad. Sci. USA* **84**:7413-7417.

Fields, S. and Song, O-K. (1989). *Nature* **340**:245-246.

Fiers, *et al.* (1978). *Nature* **273**:113.

Fincham, S.M., Hill, G.B., Hanson, J., Wijayasinghe, C. (1990). *The Prostate* **17**:189-206.

Fink, *et al.* (1992). *Hum. Gene Ther.* **3**:11-19.

Fink, D.J., *et al.* (1996). *Ann. Rev. Neurosci.* **19**:265-287.

Finkelstein, J., *et al.* (1990). *Genomics* **7**:167-172.

Fodor, S.P.A. (1997). *Science* **277**:393-395.

Freese, *et al.* (1990). *Biochem. Pharmacol.* **40**:2189-2199.

Friedman, T. (1991). *In: Therapy for Genetic Diseases*, T. Friedman, ed., Oxford University Press, pp. 105-121.

Glover, D. (1985). *DNA Cloning*, **I** and **II** (Oxford Press).

Go, R.C.P., *et al.* (1983). *J. Natl. Cancer Inst.* **71**:455-461.

Goding (1986). *Monoclonal Antibodies: Principles and Practice*, 2d ed. (Academic Press, NY).

Godowski, *et al.* (1988). *Science* **241**:812-816.

Goldgar, D.E., *et al.* (1994). *J. Natl. Can. Inst.* **86**:3:200-209.

Goldgar, D.E. et al. (1993). *Am. J. Hum. Genet.* **52**:743-748.

Gordon, *et al.* (1980). *Proc. Natl. Acad. Sci. USA* 77:7380-7384.

Gorziglia and Kapikian (1992). *J. Virol.* **66**:4407-4412.

Graham and van der Eb (1973). *Virology* **52**:456-467.

Grompe, M., (1993). *Nature Genetics* **5**:111-117.

Grompe, M., *et al.*, (1989). *Proc. Natl. Acad. Sci. USA* **86**:5855-5892.

Guthrie, G. and Fink, G.R. (1991). *Guide to Yeast Genetics and Molecular Biology* (Academic Press).

Harlow and Lane (1988). *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Hasty, P., K., *et al.* (1991). *Nature* **350**:243.

Helseth, *et al.* (1990). *J. Virol.* **64**:2416-2420.

Hodgson, J. (1991). *Bio/Technology* **9**:19-21.

Huse, *et al.* (1989). *Science* **246**:1275-1281.

Innis, *et al.* (1990). *PCR Protocols: A Guide to Methods and Applications* (Academic Press, San Diego, CA).

Jablonski, E., *et al.* (1986). *Nuc. Acids Res.* **14**:6115-6128.

Jakoby, W.B. and Pastan, I.H. (eds.) (1979). *Cell Culture. Methods in Enzymology*, Vol. 58 (Academic Press, Inc., Harcourt Brace Jovanovich (NY).

Jeffreys, *et al.* (1985). *Nature* **314**:67-73.

Johnson, *et al.* (1992). *J. Virol.* **66**:2952-2965.

Johnson, *et al.* (1993). "Peptide Turn Mimetics" *In:* Biotechnology and Pharmacy, Pezzuto et al., eds., Chapman and Hall, NY.

Kamb, A. *et al.* (1994). *Science* **264**:436-440.

Kandpal, *et al.* (1990). *Nucl. Acids Res.* **18**:1789-1795.

Kaneda, *et al.* (1989). *J. Biol. Chem.* **264**:12126-12129.

Kanehisa (1984). *Nucl. Acids Res.* **12**:203-213.

Kinszler, K.W., *et al.* (1991). *Science* **251**:1366-1370.

Kinzler, K.W. and Vogelstein, B. 1997. Gatekeepers and Caretakers. *Nature* **386**: 761-763.

Knudson, A.G. (1993). *Nature Genet.* **5**:103.

Krain, L.S. (1974). *Preventive Medicine* **3**: 154-159.

Kubo, T., *et al.* (1988). *FEBS Letts.* **241**:119.

Kyte, J. and Doolittle, R.F. (1982). *J. Mol. Biol.* **157**:105-132.

Landegren, *et al.* (1988). *Science* **242**:229.

Lathrop, G.M. et al (1985). *Am J Hum Genet.* **37**:482-489.

Lee, J.E., *et al.* (1995). *Science* **268**:836-844.

Lim, *et al.* (1992). *Circulation* **83**:2007-2011.

Lin, Y. Z. et al. (1995). *J. Biol. Chem.* **270**:14255-14258.

Litt, *et al.* (1989). *Am. J. Hum. Genet.* **44**:397-401.

Madzak, *et al.* (1992). *J. Gen. Virol.* **73**:1533-1536.

Maniatis. T., *et al.* (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Mann and Baltimore (1985). *J. Virol.* **54**:401-407.

Margolskee (1992). *Curr. Top. Microbiol. Immunol.* **158**:67-90.

Martin, R., *et al.* (1990). *BioTechniques* **9**:762-768.

Matteucci, M.D. and Caruthers, M.H. (1981). *J. Am. Chem. Soc.* **103**:3185.

Matthews and Kricka (1988). *Anal. Biochem.* **169**:1.

Meikle, A.W. et al. (1985). *Prostate* **6**:121.

Merrifield (1963). *J. Am. Chem. Soc.* **85**:2149-2156.

Metzger, *et al.* (1988). *Nature* **334**:31-36.

Miller (1992). *Curr. Top. Microbiol. Immunol.* **158**:1-24.

Miller, *et al.* (1985). *Mol. Cell. Biol.* **5**:431-437.

Miller, *et al.* (1988). *J. Virol.* **62**:4337-4345.

Mittlin (1989). *Clinical Chem.* **35**:1819.

Modrich, P. (1991). *Ann. Rev. Genet.* **25**:229-253.

Mombaerts, P., *et al.* (1992). *Cell* **68**:869.

Morganti, G. et al. (1956). *Acta Geneticae Medicae et Gemellogogiae* **6**:304-305.

Moss (1992). *Curr. Top. Microbiol. Immunol.* **158**:25-38.

Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* **158**:97-123.

Nabel (1992). *Hum. Gene Ther.* **3**:399-410.

Nabel, *et al.* (1990). *Science* **249**:1285-1288.

Nakamura, *et al.* (1987). *Science* **235**:1616-1622.

Naldini, L., *et al.* (1996). *Science* **272**:263-267.

Newton, C.R. et al (1989). *Nucl. Acids Res.* **17**:2503-2516.

Nguyen, Q., *et al.* (1992). *BioTechniques* **13**:116-123.

Novack, *et al.* (1986). *Proc. Natl. Acad. Sci. USA* **83**:586.

O'Connell, J.R. and Weeks, D.E. (1995). *Nature Genet.* **11**:402-408.

Ohi, *et al.* (1990). *Gene* **89**:279-282.

Orita, *et al.* (1989). *Proc. Natl. Acad. Sci. USA* **86**:2776-2770.

Page, *et al.* (1990). *J. Virol.* **64**:5370-5276.

Parimoo, S. et al. (1991). *Proc. Natl. Acad. Sci. USA* **88**:9623-9627.

Pellicer, *et al.* (1980). *Science* **209**:1414-1422.

Petropoulos, *et al.* (1992). *J. Virol.* **66**:3391-3397.

Philpott, K.L., *et al.* (1992). *Science* **256**:1448.

Quantin, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:2581-2584.

Rano and Kidd (1989). *Nucl. Acids Res.* **17**:8392.

Rigby, P.W.J., *et al.* (1977). *J. Mol. Biol.* **113**:237-251.

Rojas, M., Yao, S., and Lin, Y. Z. (1996). *J. Biol. Chem.* **271**:27456-27461.

Rommens, J.M. et al. (1994). *In:*"Identification of transcribed sequences", (U. Hochgeschwender and K.Gardiner, Eds.), pp. 65-79, Plenum Press, NY.

Rosenfeld, *et al.* (1992). *Cell* **68**:143-155.

Russell, D. and Hirata, R. (1998). *Nature Genetics* **18**:323-328.

Sambrook, J., *et al.* (1989), *Molecular Cloning: A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Schaffer, A.A. et al. (1994). *Human Heredity* **44**:225-237.

Scharf (1986). *Science* **233**:1076.

Schneider G, *et al.* (1998). *Nature Genetics* **18**:180-183.

Scopes, R. (1982). *Protein Purification: Principles and Practice*, (Springer-Verlag, NY).

Sheffield, V.C., *et al.* (1989). *Proc. Natl. Acad. Sci. USA* **86**:232-236.

Sheffield, V.C., *et al.* (1991). *Am. J. Hum. Genet.* **49**:699-706.

Shenk, *et al.* (1975). *Proc. Natl. Acad. Sci. USA* **72**:989.

Shimada, *et al.* (1991). *J. Clin. Invest.* **88**:1043-1047.

Shinkai, Y., *et al.* (1992). *Cell* **68**:855.

Skolnick, M.H. and Wallace, B.R. (1988). *Genomics* **2**:273-279.

Smith,J.R. et al (1996). *Science* **274**:1371.

Smith, S.A., *et al.* (1992). *Nature Genetics* **2**:128-131.

Smith, S.W. et al. (1994). The Genetic Data Environment: An expandable GUI for multiple sequence analysis. *CABIOS* **10**:671-675.

Smith, T.F. and Waterman, M.S. (1981). *J. Mol. Biol.* **147**:195-197.

Snouwaert, J.N., *et al.* (1992). *Science* **257**:1083.

Sorge, *et al.* (1984). *Mol. Cell. Biol.* **4**:1730-1737.

Spargo, C.A., *et al.* (1996). *Mol. Cell. Probes* **10**:247-256.

Stewart, *et al.* (1992). *Hum. Gene Ther.* **3**:267-275.

Stratford-Perricaudet, *et al.* (1990). *Hum. Gene Ther.* **1**:241-256.

Tavtigian, S., *et al.* (1996). *Nature Genetics* **12**:333-337.

Thierry-Mieg, D. et al. (1995). Ace.mbly. A graphic interactive program to support shotgun and directed sequencing projects.

Valancius, V. and Smithies, O. (1991). *Mol. Cell Biol.* **11**:1402.

Wagner, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3410-3414.

Wagner, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:4255-4259.

Wang and Huang (1989). *Biochemistry* **28**:9508-9514.

Wartell, R.M., *et al.* (1990). *Nucl. Acids Res.* **18**:2699-2705.

Weber and May (1989). *Am. J. Hum. Genet.* **44**:388-396.

Wells, J.A. (1991). *Methods in Enzymol.* **202**:390-411.

Wetmur and Davidson (1968). *J. Mol. Biol.* **31**:349-370.

White, M.B., *et al.*, (1992). *Genomics* **12**:301-306.

White and Lalouel (1988). *Ann. Rev. Genet.* **22**:259-279.

Wilkinson, *et al.* (1992). *Nucleic Acids Res.* **20**:2233-2239.

Wolff, *et al.* (1990). *Science* **247**:1465-1468.

Wolff, *et al.* (1991). *BioTechniques* **11**:474-485.

Woolf, C.M. (1960). *Cancer* **13**:739-744.

Wu, *et al.* (1989a). *Genomics* **4**:560-569.

Wu, *et al.* (1989b). *J. Biol. Chem.* **264**:16985-16987.

Wu, *et al.* (1991). *J. Biol. Chem.* **266**:14338-14342.

Zenke, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3655-3659.

List of Patents and Patent Applications:

U.S. Patent No. 3,817,837

U.S. Patent No. 3,850,752

U.S. Patent No. 3,939,350

U.S. Patent No. 3,996,345

U.S. Patent No. 4,275,149

U.S. Patent No. 4,277,437

U.S. Patent No. 4,366,241

U.S. Patent No. 4,376,110

U.S. Patent No. 4,486,530

U.S. Patent No. 4,554,101.

U.S. Patent No. 4,683,195

U.S. Patent No. 4,683,202

U.S. Patent No. 4,816,567

U.S. Patent No. 4,868,105

U.S. Patent No. 5,252,479

U.S. Patent No. 5,436,146.

U.S. Patent No. 5,550,050.

U.S. Patent No. 5,691,198.

U.S. Patent No. 5,735,500.

U.S. Patent No. 5,747,469.

Hitzeman *et al.*, EP 73,675A

EPO Publication No. 225,807

EP 425,731A.

European Patent Application Publication No. 0332435

Geysen, H., PCT published application WO 84/03564, published 13 September 1984

WO 90/07936.

WO 92/19195.

WO 94/25503.

WO 95/01203.

WO 95/05452.

WO 96/02286.

WO 96/02646.

WO 96/11698.

WO 96/40871.

WO 96/40959.

WO 97/12635.

PCT published application WO 93/07282

WHAT IS CLAIMED IS:

1. An isolated nucleic acid selected from the group consisting of a nucleic acid having SEQ ID NO:1, a nucleic acid having SEQ ID NO:3, a nucleic acid having SEQ ID NO:5, a nucleic acid having SEQ ID NO:7, a nucleic acid having SEQ ID NO:8, a nucleic acid having SEQ ID NO:10, a nucleic acid having SEQ ID NO:12, a nucleic acid having SEQ ID NO:14, a nucleic acid having SEQ ID NO:16, a nucleic acid having SEQ ID NO:18, a nucleic acid having SEQ ID NO:20, a nucleic acid having SEQ ID NO:22, a nucleic acid having SEQ ID NO:24, a nucleic acid having SEQ ID NO:26, a nucleic acid having SEQ ID NO:28, a nucleic acid having SEQ ID NO:30, a nucleic acid having SEQ ID NO:32, a nucleic acid having SEQ ID NO:34, a nucleic acid having SEQ ID NO:36, a nucleic acid having SEQ ID NO:38, a nucleic acid having SEQ ID NO:39, a nucleic acid having SEQ ID NO:41, a nucleic acid having SEQ ID NO:43, a nucleic acid having SEQ ID NO:45, a nucleic acid having SEQ ID NO:47, a nucleic acid having SEQ ID NO:49, and a nucleic acid having SEQ ID NO:51.

2. An isolated nucleic acid selected from the group consisting of a nucleic acid having SEQ ID NO:2, a nucleic acid having SEQ ID NO:4, a nucleic acid having SEQ ID NO:6, a nucleic acid having SEQ ID NO:9, a nucleic acid having SEQ ID NO:11, a nucleic acid having SEQ ID NO:13, a nucleic acid having SEQ ID NO:15, a nucleic acid having SEQ ID NO:17, a nucleic acid having SEQ ID NO:19, a nucleic acid having SEQ ID NO:21, a nucleic acid having SEQ ID NO:23, a nucleic acid having SEQ ID NO:25, a nucleic acid having SEQ ID NO:27, a nucleic acid having SEQ ID NO:29, a nucleic acid having SEQ ID NO:31, a nucleic acid having SEQ ID NO:33, a nucleic acid having SEQ ID NO:35, a nucleic acid having SEQ ID NO:37, a nucleic acid having SEQ ID NO:40, a nucleic acid having SEQ ID NO:42, a nucleic acid having SEQ ID NO:44, a nucleic acid having SEQ ID NO:46, a nucleic acid having SEQ ID NO:48, and a nucleic acid having SEQ ID NO:52.

3. An isolated DNA coding for a wild-type HPC1 polypeptide or a portion of an HPC1 polypeptide wherein said portion is encoded by any one or more of the nucleic acids selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 39, 41, 43, 45, 47, 49 and 51.

4. An isolated DNA encoding a variant of a wild-type HPC1 polypeptide or a portion of an HPC1 polypeptide wherein said DNA is encoded by any one or more allelic variants of the nucleic acids selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 39, 41, 43, 45, 47, 49 and 51.

5. The isolated DNA of claim 4 wherein said allelic variant is selected from the group consisting of the variants set forth in Tables 11 and 13.

6. The isolated DNA of any one of claims 3 to 5 which contains HPC1 regulatory sequences.

7. An isolated DNA having at least 15 consecutive nucleotides of the isolated DNA of any one of claims 1 to 5.

8. An isolated DNA coding for a mutated form of HPC1 polypeptide.

9 The isolated DNA of claim 8, wherein the DNA comprises a mutated form of any one or more of the nucleotide sequences set forth in SEQ ID NOs:1-52.

10. The isolated DNA of claim 8, wherein said mutated form is selected from the group consisting of the mutations set forth in Table 10.

11. The isolated DNA of claim 9, wherein the mutation is selected from the group consisting of a deletion mutation, a nonsense mutation, an insertion mutation and a missense mutation.

12. An isolated DNA having at least 15 consecutive nucleotides of the DNA of any one of claims 8-11.

13. The isolated DNA of claim 12, wherein the isolated DNA overlaps the mutation.

14. An isolated DNA selected from the group consisting of:
(1) a DNA having the nucleotide sequence set forth in SEQ ID NO:47 having an A at nucleotide position 42,

(2) a DNA having the nucleotide sequence set forth in SEQ ID NO:38 having a T at nucleotide position 31,

(3) a DNA having the nucleotide sequence set forth in SEQ ID NO:39 having a T at nucleotide position 31,

(4) a DNA having the nucleotide sequence set forth in SEQ ID NO:4 having an A at nucleotide position 207, and

(5) a DNA having the nucleotide sequence set forth in SEQ ID NO:20 having a T at base 158.

15. A replicative cloning vector which comprises the isolated DNA of any one of claims 1 to 5 and 8 to 10 or parts thereof and a replicon operative in a host cell.

16. An expression system which comprises the isolated DNA of any one of claims 1 to 5 and 8 to 10 operably linked to suitable control sequences.

17. Recombinant host cells transformed with the expression system of claim 15.

18. A method of producing recombinant HPC1 polypeptide which comprises culturing the cells of claim 17 under conditions effective for the production of said HPC1 polypeptide and harvesting the recombinant HPC1 polypeptide.

19. A preparation of human HPC1 polypeptide substantially free of other human proteins.

20. A preparation of human polypeptide substantially free of other human proteins, the amino acid sequence of said polypeptide having substantial sequence homology with a wild-type HPC1 polypeptide, and said human polypeptide having substantially similar function as a wild-type HPC1 polypeptide.

21. An antibody immunoreactive with a human HPC1 polypeptide or portion thereof.

22. The antibody of claim 21 which is a polyclonal antibody.

23. The antibody of claim 21 which is a monoclonal antibody.

24. A pair of single-stranded DNA primers for determination of a nucleotide sequence of an HPC1 gene by an amplification reaction, the sequence of said primers being derived from human chromosome 1, wherein the use of said primers in an amplification results in the synthesis of DNA having all or part of the sequence of the HPC1 gene.

25. The pair of primers of claim 24 wherein said HPC1 gene has a nucleotide sequence set forth in any one or more of SEQ ID NOs:1-52.

26. A method for identifying a mutant HPC1 nucleotide sequence in a suspected mutant HPC1 allele which comprises comparing the nucleotide sequence of the suspected mutant HPC1 allele with a wild-type HPC1 nucleotide sequence, wherein a difference between the suspected mutant and a wild-type sequences identifies a mutant HPC1 nucleotide sequence.

27. A method for identifying a polymorphic HPC1 nucleotide sequence in a suspected polymorphic HPC1 allele which comprises comparing the nucleotide sequence of the suspected polymorphic HPC1 allele with a wild-type HPC1 nucleotide sequence of the said HPC1 gene having a nucleotide sequence set forth in any one or more of SEQ ID NOs:1-31, wherein a difference between the suspected polymorphic and wild-type sequences identifies a polymorphic HPC1 nucleotide sequence.

28. A method for identifying a consensus HPC1 nucleotide sequence which comprises analyzing the sequence of at least 5 individuals having a wild-type allele and identifying any polymorphisms and their frequency in the population examined.

29. A kit for detecting mutations in the HPC1 gene resulting in a susceptibility to prostate cancer which comprises at least one oligonucleotide primer specific for an HPC1 gene mutation and instructions relating to detecting mutations in the HPC1 gene.

30. A kit for detecting mutations in the HPC1 gene resulting in a susceptibility to prostate cancer which comprises at least one allele-specific oligonucleotide probe for an HPC1 gene mutation and instructions relating to detecting mutations in the HPC1 gene.

31. A method for supplying a wild-type HPC1 gene function or an HPC1 function substantially similar to the wild-type to a cell which has lost said gene function or has altered gene function by virtue of a mutation in a HPC1 gene, wherein said method comprises: introducing into the cell a nucleic acid which suppresses a transformed state of said cell, said nucleic acid selected from the group consisting of a wild-type HPC1 gene nucleic acid, a portion of said wild-type HPC1 gene nucleic acid, a nucleic acid substantially homologous and has substantially similar function to said wild-type HPC1 gene nucleic acid and a portion of said nucleic acid substantially homologous to said wild-type HPC1 gene nucleic acid.

32. The method of claim 31 wherein said nucleic acid is a wild-type HPC1 gene nucleic acid.

33. The method of claim 31 or 32 wherein said nucleic acid contains an HPC1 gene regulatory sequences.

34. The method of any one of claims 31 to 33 wherein said nucleic acid is incorporated into the genome of said cell.

35. A method for supplying a wild-type HPC1 gene function or an HPC1 function substantially similar to the wild-type to a cell which has lost said gene function or has altered gene function by virtue of a mutation in the HPC1 gene, comprising: introducing into the cell a molecule which suppresses a transformed state of said cell, said molecule selected from the group consisting of a wild-type HPC1 polypeptide, a portion of said wild-type HPC1 polypeptide, a polypeptide substantially homologous to said wild-type HPC1 polypeptide, a portion of said polypeptide substantially homologous to said wild-type HPC1 polypeptide and a molecule which mimics the function of said wild-type HPC1 polypeptide.

36. The method of claim 35 wherein said molecule is a wild-type HPC1 polypeptide.

37. A method for screening potential cancer therapeutics which comprises: combining (i) an HPC1 binding partner, (ii) an HPC1 polypeptide having a portion of said amino acid sequence which binds to said binding partner and (iii) a compound suspected of being a cancer therapeutic and determining the amount of binding of the HPC1 polypeptide to its binding partner.

38. A method for screening potential cancer therapeutics which comprises: combining an HPC1 binding partner and a compound suspected of being a cancer therapeutic and measuring the biological activity of the binding partner.

39. A method for screening potential cancer therapeutics, wherein said method comprises: growing a transformed eukaryotic host cell containing an altered HPC1 gene causing cancer in the presence of a compound suspected of being a cancer therapeutic, growing said transformed eukaryotic host cell in the absence of said compound, determining the rate of growth of said host cell in the presence of said compound and the rate of growth of said host cell in the absence of said compound and comparing the growth rate of said host cells, wherein a slower rate of growth of said host cell in the presence of said compound is indicative of a cancer therapeutic.

40. A method for screening potential cancer therapeutics which comprises: administering a compound suspected of being a cancer therapeutic to a transgenic animal which carries an altered HPC1 allele from a second animal in its genome and determining the development or growth of a cancer lesion.

41. A transgenic animal which carries an altered HPC1 allele.

42. The transgenic animal of claim 41 wherein the altered HPC1 allele is selected from the group consisting of a deletion mutation, a nonsense mutation, a frameshift mutation and a missense mutation.

43. The transgenic animal of claim 41 wherein the altered HPC1 allele is a disrupted allele.

44. A method for screening germline of a human subject for an alteration of an HPC1 gene, wherein said method comprises comparing germline sequence of an HPC1 gene or HPC1 RNA from a tissue sample from said subject or a sequence of HPC1 cDNA made from mRNA from said sample or an HPC1 polypeptide isolated from said sample with germline sequences of wild-type HPC1 gene, wild-type HPC1 RNA, wild-type HPC1 cDNA or a wild-type HPC1 polypeptide wherein a difference in the sequence of the HPC1 gene, HPC1 RNA, HPC1 cDNA or HPC1 polypeptide of the subject from wild-type indicates an alteration in the HPC1 gene in said subject.

45. The method of claim 44 wherein said wild-type HPC1 gene has a nucleotide sequence set forth in any one or more of SEQ ID NOs:1-52.

46. The method of claim 44 wherein the nucleic acid sequence of HPC1 RNA from the subject is compared to nucleic acid sequences of wild-type HPC1 gene, HPC1 RNA or HPC1 cDNA.

47. The method of claim 44 wherein the nucleic acid sequence is compared by hybridizing an HPC1 gene probe which specifically hybridizes to an HPC1 allele to RNA isolated from said subject and detecting of the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

48. The method of claim 44 wherein the HPC1 polypeptide from the subject is compared to a wild-type HPC1 polypeptide.

49. The method of claim 44 wherein a regulatory region of the HPC1 gene from said subject is compared with a regulatory region of wild-type HPC1 gene sequences.

50. The method of claim 44 wherein a germline nucleic acid sequence is compared by obtaining a first HPC1 gene fragment from an HPC1 gene from a human sample and a second HPC1 gene fragment from a wild-type HPC1 gene, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first HPC1

gene fragment and from said second HPC1 gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first HPC1 gene fragment is shifted relative to said second HPC1 gene fragment and sequencing said single-stranded DNA from said first HPC1 gene fragment having a shift in electrophoretic mobility.

51. The method of claim 44 wherein a germline nucleic acid sequence is compared by hybridizing an HPC1 gene probe which specifically hybridizes to an HPC1 allele to genomic DNA isolated from said sample and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

52. The method of claim 44 wherein the alteration in the germline sequence is detected by hybridization with an allele-specific probe.

53. The method of claim 44 wherein a germline nucleic acid sequence is compared by amplifying all or part of an HPC1 gene from said sample using a set of primers to produce amplified nucleic acids and sequencing said amplified nucleic acids.

54. The method of claim 44 wherein a germline nucleic acid sequence is compared by amplifying all or part of an HPC1 gene using a primer specific for a specific HPC1 mutant allele and detecting the presence of an amplified product, wherein a presence of said product indicates the presence of said specific allele.

55. The method of claim 44 wherein a germline nucleic acid sequence is compared by molecularly cloning all or part of said HPC1 gene from said sample to produce cloned nucleic acid and sequencing the cloned nucleic acid.

56. The method of claim 44 wherein a germline nucleic acid sequence is compared by obtaining a first HPC1 gene fragment from (a) HPC1 gene genomic DNA isolated from said sample, (b) HPC1 RNA isolated from said sample or (c) HPC1 cDNA made from

mRNA isolated from said sample, obtaining a second HPC1 gene fragment from (a) wild-type HPC1 genomic DNA, (b) wild-type HPC1 RNA or (c) wild-type cDNA made from wild-type mRNA, said second HPC1 gene fragment corresponding to said first HPC1 gene fragment, forming single-stranded DNA from said first HPC1 gene fragment and from said second HPC1 gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first HPC1 gene fragment and single-stranded DNA from said second HPC1 gene fragment, analyzing the heteroduplex to determine if said single-stranded DNA from said first HPC1 gene fragment has a mismatch relative to said single-stranded DNA from said second HPC1 gene fragment and sequencing said first single-stranded DNA from said first HPC1 gene fragment having a mismatch.

57. The method of claim 44 wherein a germline nucleic acid sequence is compared by amplifying HPC1 nucleic acids from said sample to produce amplified nucleic acids, hybridizing the amplified nucleic acids to an HPC1 DNA probe specific for an HPC1 allele and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the subject.

58. The method of claim 44 wherein the alteration in the germline sequence is detected by amplification of HPC1 gene sequences in said tissue and hybridization of the amplified sequences to nucleic acid probes which comprise mutant HPC1 gene sequences.

59. The method of claim 44 wherein a germline nucleic acid sequence is compared by analyzing HPC1 nucleic acids in said sample for a mutation from the group consisting of a deletion mutation, a point mutation and an insertion mutation.

60. The method of claim 44 wherein a germline nucleic acid sequence is compared by hybridizing the tissue sample *in situ* with a nucleic acid probe specific for an HPC1 allele and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

61. The method of claim 48 wherein the alteration in the germline sequence is detected by immunoblotting.

62. The method of claim 48 wherein the alteration in the germline sequence is detected by immunocytochemistry.

63. The method of claim 48 wherein the alteration in the germline sequence is detected by assaying for binding interactions between HPC1 gene protein isolated from said tissue and its ligand.

64. The method of claim 48 wherein the alteration in the germline sequence is detected by assaying for the inhibition of biochemical activity of its binding partner.

65. The method of claim 48, wherein an HPC1 polypeptide from a tissue sample from said subject is analyzed for an altered HPC1 polypeptide by (i) detecting either a full length HPC1 polypeptide or a truncated HPC1 polypeptide or (ii) contacting an antibody which specifically binds to an epitope of an altered HPC1 polypeptide to the HPC1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of a germline alteration in the HPC1 gene.

66. The method of claim 65 wherein an HPC1 polypeptide is analyzed by detecting a truncated HPC1 polypeptide.

67. The method of claim 65 wherein an HPC1 polypeptide is analyzed by contacting an antibody which specifically binds to an epitope of an altered HPC1 polypeptide to the HPC1 polypeptide from said sample.

68. A kit for screening for an alteration in an HPC1 gene in a human subject which comprises at least one antibody (i) which specifically binds to wild-type HPC1 polypeptide but not a truncated HPC1 polypeptide or (ii) which specifically binds to an epitope of an altered HPC1 polypeptide.

69. A method for screening a tumor sample from a human subject for the presence of a somatic alteration in an HPC1 gene in said tumor which comprises comparing a sequence of an HPC1 gene from said tumor sample, or HPC1 RNA from said tumor sample, or a sequence of HPC1 cDNA made from mRNA from said sample or an HPC1 polypeptide isolated from said sample with sequences of wild-type HPC1 gene, wild-type HPC1 RNA, wild-type HPC1 cDNA or a wild-type HPC1 polypeptide from a nontumor sample from said subject, wherein a difference in the sequence of the HPC1 gene, HPC1 RNA, HPC1 cDNA or HPC1 polypeptide of the tumor sample from the nontumor sample indicates an alteration in the HPC1 gene in said tumor sample.

70. The method of claim 69 wherein said wild-type HPC1 gene has a nucleotide sequence set forth in any one or more of SEQ ID NOs:1-31.

71. The method of claim 69 wherein the nucleic acid sequence of HPC1 RNA from the subject is compared to nucleic acid sequences of wild-type HPC1 gene, HPC1 RNA or HPC1 cDNA.

72. The method of claim 69 wherein the nucleic acid sequence is compared by hybridizing an HPC1 gene probe which specifically hybridizes to an HPC1 allele to RNA isolated from said tumor sample and detecting of the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the tumor sample.

73. The method of claim 69 wherein the HPC1 polypeptide from the tumor sample is compared to the HPC1 polypeptide from the nontumor sample.

74. The method of claim 69 wherein a regulatory region of the HPC1 gene from said tumor sample is compared with a regulatory region of the HPC1 gene from the nontumor sample.

75. The method of claim 69 wherein the nucleic acid sequence is compared by obtaining a first HPC1 gene fragment from an HPC1 gene from the tumor sample and a second HPC1 gene fragment from the nontumor sample, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first HPC1 gene fragment and from said

second HPC1 gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first HPC1 gene fragment is shifted relative to said second HPC1 gene fragment and sequencing said single-stranded DNA from said first HPC1 gene fragment having a shift in electrophoretic mobility.

76. The method of claim 69 wherein a nucleic acid sequence is compared by hybridizing an HPC1 gene probe which specifically hybridizes to an HPC1 allele to genomic DNA isolated from said tumor sample and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the tumor.

77. The method of claim 69 wherein the alteration in the nucleic acid sequence is detected by hybridization with an allele-specific probe.

78. The method of claim 69 wherein a nucleic acid sequence is compared by amplifying all or part of an HPC1 gene from said tumor sample using a set of primers to produce amplified nucleic acids and sequencing said amplified nucleic acids.

79. The method of claim 69 wherein a nucleic acid sequence is compared by amplifying all or part of an HPC1 gene in the tumor sample using a primer specific for a specific HPC1 mutant allele and detecting the presence of an amplified product, wherein a presence of said product indicates the presence of said specific allele.

80. The method of claim 69 wherein a nucleic acid sequence is compared by molecularly cloning all or part of said HPC1 gene from said tumor sample to produce cloned nucleic acid and sequencing the cloned nucleic acid.

81. The method of claim 69 wherein a nucleic acid sequence is compared by obtaining a first HPC1 gene fragment from (a) HPC1 gene genomic DNA isolated from said tumor sample, (b) HPC1 RNA isolated from said tumor sample or (c) HPC1 cDNA made from mRNA isolated from said tumor sample, obtaining a second HPC1 gene fragment from (a) HPC1 genomic DNA from said nontumor sample, (b) HPC1 RNA from said nontumor sample or

(c) cDNA made from said mRNA from said nontumor sample, said second HPC1 gene fragment corresponding to said first HPC1 gene fragment, forming single-stranded DNA from said first HPC1 gene fragment and from said second HPC1 gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first HPC1 gene fragment and single-stranded DNA from said second HPC1 gene fragment, analyzing the heteroduplex to determine if said single-stranded DNA from said first HPC1 gene fragment has a mismatch relative to said single-stranded DNA from said second HPC1 gene fragment and sequencing said first single-stranded DNA from said first HPC1 gene fragment having a mismatch.

82. The method of claim 69 wherein a nucleic acid sequence is compared by amplifying HPC1 nucleic acids from said tumor sample to produce amplified nucleic acids, hybridizing the amplified nucleic acids to an HPC1 DNA probe specific for an HPC1 allele and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the tumor sample.

83. The method of claim 69 wherein the alteration in the sequence is detected by amplification of HPC1 gene sequences in said tumor sample and hybridization of the amplified sequences to nucleic acid probes which comprise mutant HPC1 gene sequences.

84. The method of claim 69 wherein a nucleic acid sequence is compared by analyzing HPC1 nucleic acids in said sample for a mutation from the group consisting of a deletion mutation, a point mutation and an insertion mutation.

85. The method of claim 69 wherein a nucleic acid sequence is compared by hybridizing the tissue sample *in situ* with a nucleic acid probe specific for an HPC1 allele and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the tumor sample.

86. The method of claim 73 wherein the alteration in the sequence is detected by immunoblotting.

87. The method of claim 73 wherein the alteration in the sequence is detected by immunocytochemistry.

88. The method of claim 73 wherein the alteration in the sequence is detected by assaying for binding interactions between HPC1 gene protein isolated from said tumor sample and its ligand.

89. The method of claim 73 wherein the alteration in the sequence is detected by assaying for the inhibition of biochemical activity of its binding partner.

90. The method of claim 73, wherein an HPC1 polypeptide from a tumor sample from said subject is analyzed for an altered HPC1 polypeptide by (i) detecting either a full length HPC1 polypeptide or a truncated HPC1 polypeptide or (ii) contacting an antibody which specifically binds to an epitope of an altered HPC1 polypeptide to the HPC1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of an alteration in the HPC1 gene.

91. The method of claim 90 wherein an HPC1 polypeptide is analyzed by detecting a truncated HPC1 polypeptide.

92. The method of claim 90 wherein an HPC1 polypeptide is analyzed by contacting an antibody which specifically binds to an epitope of an altered HPC1 polypeptide to the HPC1 polypeptide from said sample.

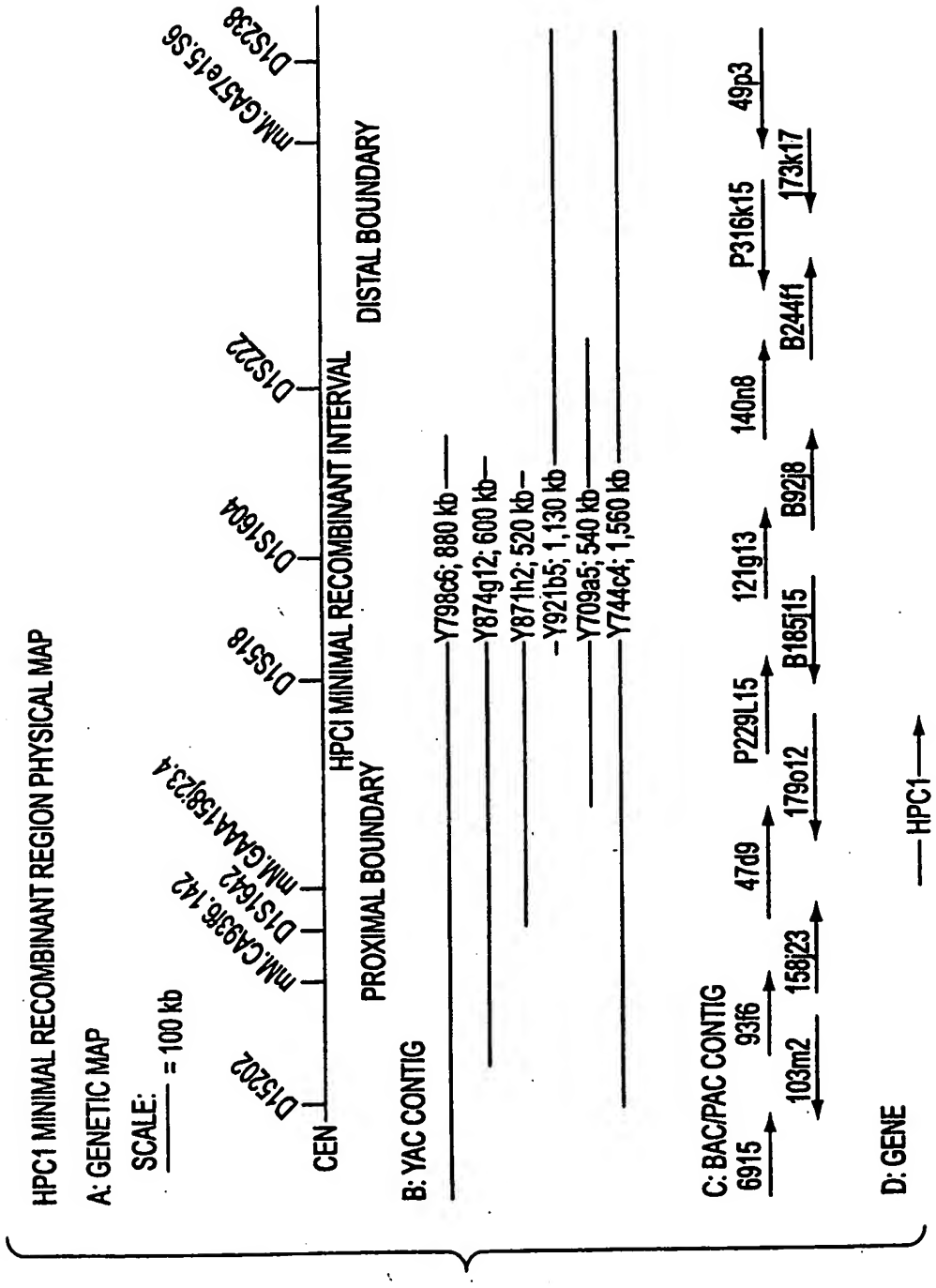

FIG. 1

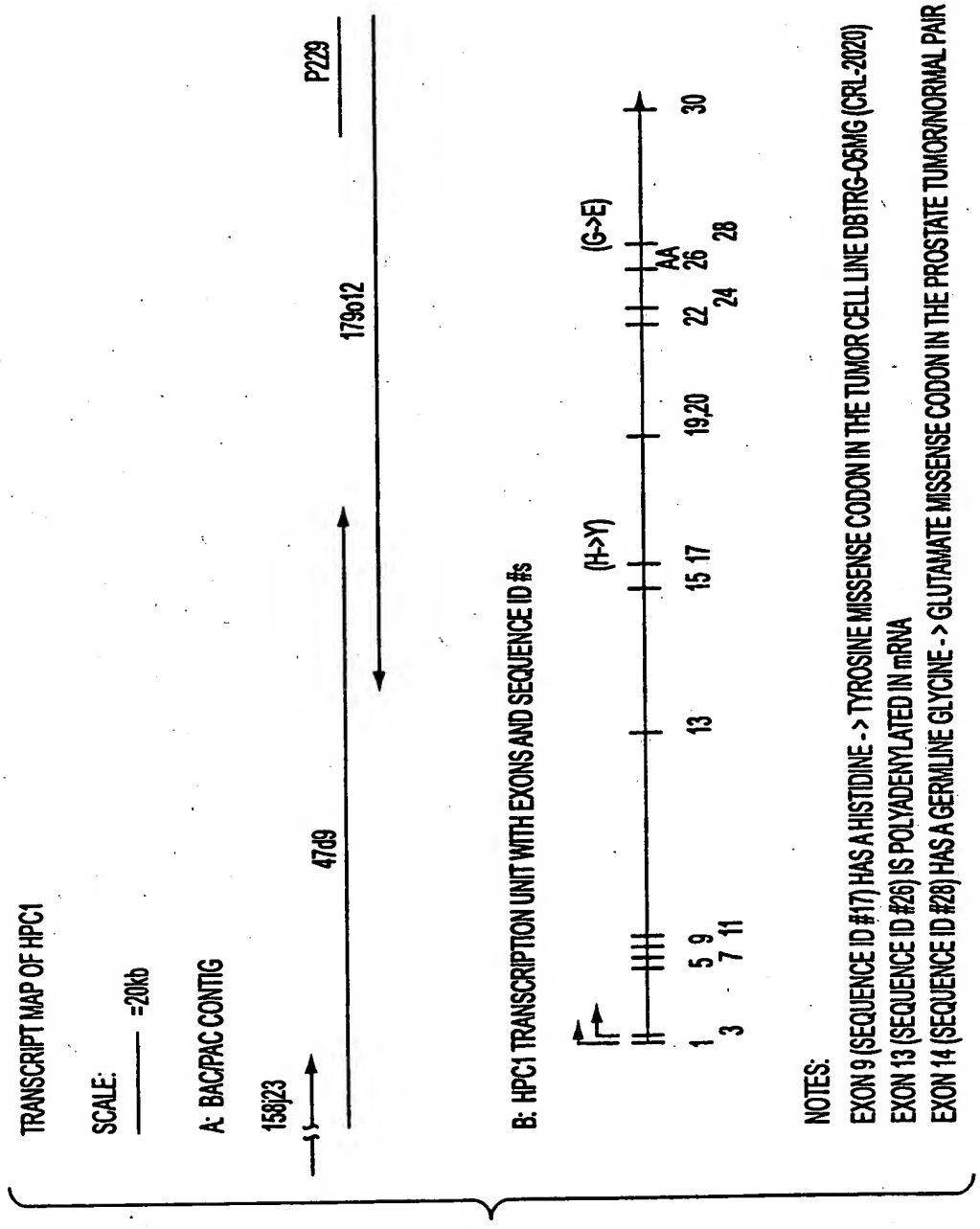

FIG. 2

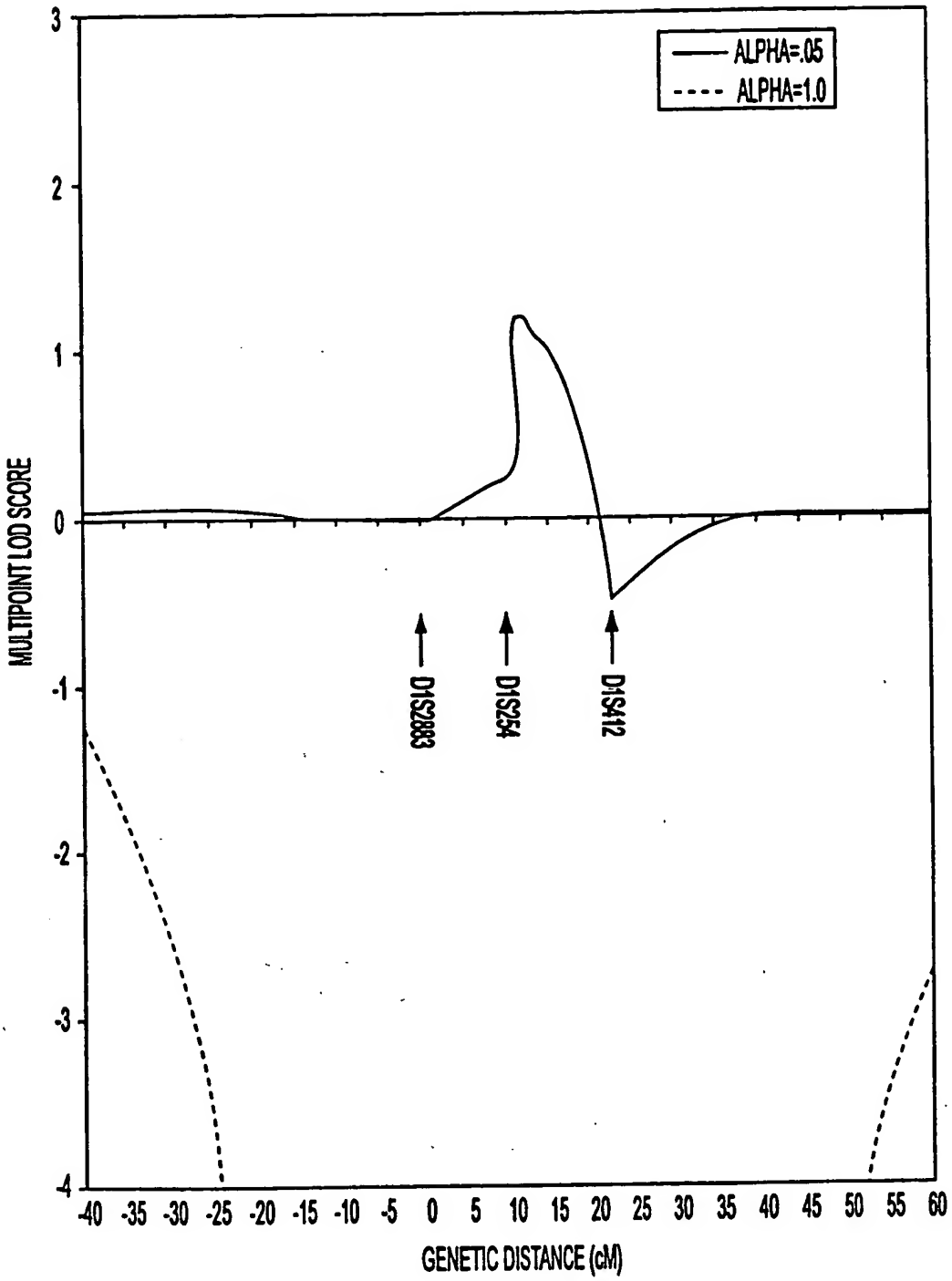

FIG. 3

SEQUENCE LISTING

<110> Myriad Genetics, Inc.

<120>

<130>

<140>
<141>

<160>

<170>

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 74 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g1m20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
 1    GTGCTCAAAAAACATTTGTTGAGTAAGTGAACCTGAGACTATCAACAAGC
51    ATTATTTTAAAATCACTAGCAAAG
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME:g1m20.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:141-214

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
  1   tggtattttt ggttgattta ttaatttctt ttctgtcttt ccactctaat
 51   ctcaaacata aatgtcccga agcagaaacc ctatacaact tcttcagtac
101   ttttcctgcc agtgcttaaa aggatgcctt tttattgtag GTGCTCAAAA
151   AACATTTGTT GAGTAAGTGA ACCTGAGACT ATCAACAAGC ATTATTTTAA
201   AATCACTAGC AAAGgtaagt aagt
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 161 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g2ml3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
  1  AAAAAAAGCC ATTTTCCTCT CTCCACCACA TACCACACTT CTGGGATTCT
 51  GAAATATCTT ACCGAGCACT TTATTCAATC TAAATTTAAA TAGAAGTTTT
101  CCCCACTTCC CAAGAGAGAA ACAACAACGA ACTAGATGAG AATGAGAGGA
151  ACTGGAAGAA G
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 253 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g2ml3.genomic

(ix) FEATURE:
(A) NAME/KEY:exon: 40-200

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
  1  ctcacaggcc aacgtattca aaagctccag gaaaaaaaaA AAAAAAGCCA
 51  TTTTCCTCTC TCCACCACAT ACCACACTTC TGGGATTCTG AAATATCTTA
101  CCGAGCACTT TATTCAATCT AAATTTAAAT AGAAGTTTTC CCCACTTCCC
151  AAGAGAGAAA CAACAACGAA CTAGATGAGA ATGAGAGGAA CTGGAAGAAG
201  gtaatgctcc atatcatttg gttaatctat tcttgtttat taatttatta
251  cat
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g3ml

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
  1  TAATGAAATC TGAGAAGCTG AATTTAGCAA TACAGATGCA AACTGTGCCA
 51  TCAGAAGATT AAAATGAAAG TGAAATGTCC TGAAAATATC AGAATCGCAT
101  CAGTAATAGA AGTAAATGAA AAGTGAAGAC CTCTTTGAAT TATCTTATTT
151  CATTTGACTA TGTTCCTCCT GAGTCACAAA AAAAGGATGT TACAGCTATT
201  TTTCTTAAGC TGATGGGCCA AAAGATTGTA
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g3m1.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 51-280

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
  1  taattcatat gaagaaaagt tttttgtatt tgtttttttgg taaaatgcag
 51  TAATGAAATC TGAGAAGCTG AATTTAGCAA TACAGATGCA AACTGTGCCA
101  TCAGAAGATT AAAATGAAAG TGAAATGTCC TGAAAATATC AGAATCGCAT
151  CAGTAATAGA AGTAAATGAA AAGTGAAGAC CTCTTTGAAT TATCTTATTT
201  CATTTGACTA TGTTCCTCCT GAGTCACAAA AAAAGGATGT TACAGCTATT
251  TTTCTTAAGC TGATGGGCCA AAAGATTGTA gtaagtattt tgacgattgt
301  ctggtggggg
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 977 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g4m17a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
  1  TTCTAGGCTC CAGGGACACA AGAGTGACCA AAACTGACAG AACAGGTCAT
 51  CATGGAGCTT TCAATGTGGC TGAGGATACC TGAGGAAAAA AACATCTGCT
101  GTCAACCTAA ACTAAACTTT GCTATCAACC ACAATAGGAT AACACAGGGT
151  GCTGAGAGAA TAACACAGAG GTGCTTGAGT AACATGAGCT CCTAACTGAC
201  CACCACTAAG TATGTGGTTC TACCAAATGA ATAAAATCAC TGGCCATCCT
251  ATTAGTTTTG TCTTTGCTTA GCCTAATGAT AGGAAACGAA TGGAACTTTG
301  CCATCAAACG AATACAAGAT AATCCCTATT GCTTGAATTC TTAATTGTTT
351  TTTGGACTAC CGATATATTT TACTTCATTT GGTACCGTAG TTTACATATT
401  TTAAATAAAG TATTATACAT AATATCTTAC ATAAATATGA ATGATTTCAA
451  GGTGCTACAT CCCTGGAATT GTTATATTTT AATCTGTTGA TGTTGCTAAA
501  ACCCACAGTG GTGGCAATGT TACTAAAATT GAATTCATTC TTTCACTTGA
551  AGGTAATGCA TAAAGAGGTA AGGCCAAAAT ACTGAACGAA ACTTAATATA
601  GTAATTAAAG TGTGCTGTAA ATTTTGCACT TCTTATTAAA GTTTAGTTTC
651  AATTAATTAG CATCTTGGAA ATAAAAAGGA TAATTTTTAA AGTATTCTAA
701  TTTTCAATAA ATAAAAAGAA AATATTACTA AGATTCCTAG GATATATTGA
751  TCAATACTAT CCATTAATGT AACTGAAAGT GGTTAGAAAA TTTCAGGACC
801  AACTACTGTA AACTAAAATC AGAGCTTTAG TTCATCCATA GCCAAAAATA
851  TTTATCTCTG TGTTCTGCAG AGCATAAAGC TAGGGAAGTA AATAAGCATT
901  GTGTGAAATT TGTGATTAAA ATGAACTTCT ATTTAAATAA AAAAGAAAAA
951  GCACTTGATA AGaaaaaaaa aaaaaaa
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g4m17b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
  1  TTCTAGGCTC CAGGGACACA AGAGTGACCA AAACTGACAG AACAGGTCAT
 51  CATGGAGCTT TCAATGTGGC TGAGGATACC TGAGGAAAAA AACATCTGCT
101  GTCAACCTAA ACTAAACTTT GCTATCAACC ACAATAGGAT AACACAGGGT
151  GCTGAGAGAA TAACACAGAG
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1090 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g4m17.genomic

(ix) FEATURE:
(A) NAME/KEY:ALTERNATE EXON (a): 57-1018
(A) NAME/KEY:ALTERNATE EXON (b): 57-226

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
   1  atttattcag ttatttaaca aaagctaatt gaatgcctgc tgtatttctg
  51  atgcagTTCT AGGCTCCAGG GACACAAGAG TGACCAAAAC TGACAGAACA
 101  GGTCATCATG GAGCTTTCAA TGTGGCTGAG GATACCTGAG GAAAAAAACA
 151  TCTGCTGTCA ACCTAAACTA AACTTTGCTA TCAACCACAA TAGGATAACA
 201  CAGGGTGCTG AGAGAATAAC ACAGAGGTGC TTGAGTAACA TGAGCTCCTA
 251  ACTGACCACC ACTAAGTATG TGGTTCTACC AAATGAATAA AATCACTGGC
 301  CATCCTATTA GTTTTGTCTT TGCTTAGCCT AATGATAGGA AACGAATGGA
 351  ACTTTGCCAT CAAACGAATA CAAGATAATC CCTATTGCTT GAATTCTTAA
 401  TTGTTTTTTG GACTACCGAT ATATTTTACT TCATTTGGTA CCGTAGTTTA
 451  CATATTTTAA ATAAAGTATT ATACATAATA TCTTACATAA ATATGAATGA
 501  TTTCAAGGTG CTACATCCCT GGAATTGTTA TATTTTAATC TGTTGATGTT
 551  GCTAAAACCC ACAGTGGTGG CAATGTTACT AAAATTGAAT TCATTCTTTC
 601  ACTTGAAGGT AATGCATAAA GAGGTAAGGC CAAAATACTG AACGAAACTT
 651  AATATAGTAA TTAAAGTGTG CTGTAAATTT TGCACTTCTT ATTAAAGTTT
 701  AGTTTCAATT AATTAGCATC TTGGAAATAA AAAGGATAAT TTTTAAAGTA
 751  TTCTAATTTT CAATAAATAA AAAGAAAATA TTACTAAGAT TCCTAGGATA
 801  TATTGATCAA TACTATCCAT TAATGTAACT GAAAGTGGTT AGAAAATTTC
 851  AGGACCAACT ACTGTAAACT AAAATCAGAG CTTTAGTTCA TCCATAGCCA
 901  AAAATATTTA TCTCTGTGTT CTGCAGAGCA TAAAGCTAGG GAAGTAAATA
 951  AGCATTGTGT GAAATTTGTG ATTAAAATGA ACTTCTATTT AAATAAAAAA
1001  GAAAAAGCAC TTGATAAGac agaatgtctc catgatgttc cctccacgtt
1051  tatctatatt tatatcaact atatctgtct atattatctg
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 220 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g5m19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
  1  AAAAATAACC TGATACAAGC AGAATTGTCA AGGAATGTCA TATGTCAAAA
 51  AATGAAGTGA CAAATGAGAA AGGGCTGAGT CACTTAGAAA ATATCTTGCA
101  GACAGCCAAC GGAAACAAGC ACTAAATTAA AGAGAGTGGC TGGTGGAAAT
151  AGGTCAAAAC AGCCCAGTGA AGCATAGTCG TTATATGACA AAATGACCAG
201  AAGATGTGTT TACATTACTG
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g5m19.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:65-284

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
  1  ttttaaactt gtttctcttt cttttattaa atacaaccct gtattggcat
 51  tttccatttg atagAAAAAT AACCTGATAC AAGCAGAATT GTCAAGGAAT
101  GTCATATGTC AAAAAATGAA GTGACAAATG AGAAAGGGCT GAGTCACTTA
151  GAAAATATCT TGCAGACAGC CAACGGAAAC AAGCACTAAA TTAAAGAGAG
201  TGGCTGGTGG AAATAGGTCA AAACAGCCCA GTGAAGCATA GTCGTTATAT
251  GACAAAATGA CCAGAAGATG TGTTTACATT ACTGgtatgt aaactacata
301  ttatcagagg cttgatattg gtgaccttca aattaagtac cttaattttg
351  tcggat
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g6m18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
  1  GTGGTCAAGA AAAGTTTCAT TGAGAAGATA CAATTTGGTC AGAGACTTGA
 51  TAAGGTGAAG CTCATCCTGT TGATATCACC CTGTGGAGAA ACAGCATTTC
```

101 CAGGAGGTAA AAATA

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 286 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g6m18.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:80-194

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
  1  ctactgcaaa aagtaataaa ataaagaaga aaagggagat aagtagtgcc
 51  attaggtaat gactttcatt ttaaatgagG TGGTCAAGAA AAGTTTCATT
101  GAGAAGATAC AATTTGGTCA GAGACTTGAT AAGGTGAAGC TCATCCTGTT
151  GATATCACCC TGTGGAGAAA CAGCATTTCC AGGAGGTAAA AATAgtaagt
201  acaagaggct cctagacgga caagcttggt attcttgaga gagcaaggag
251  accagtgtgg ctgctatgta cagtggccaa ggggaa
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g7m10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
  1  ATGGAGTTGT CTCACTCTGG CACGATCTCA GCTCACTGCA ACCTCCGCCT
101  CCTGGGTTCG AGAGATTCTC CTGCCTCACT CAGCCTCCCA G
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g7m10.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON:33-181

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
   1  atttgttttt cctataccca cttccaaagt agTATGTATT TTCTTTCGCC
  51  ACACTGTTTA CATTTATTTA TTTGTTTATT TATTTTTGAG ATGGAGTTGT
 101  CTCACTCTGG CACGATCTCA GCTCACTGCA ACCTCCGCCT CCTGGGTTCG
 151  AGAGATTCTC CTGCCTCACT CAGCCTCCCA Ggtagccggg attacaggca
 201  tgcaccacca tgcc
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g8m11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
   1  ACATAGCTGG AAGGCACCAT CCATGAACCA ACAAATAGGC TCTCACCAGA
  51  AATCAAATAT TCCTTGATCT TGGATTTTTC AGCCTCCAGA ACT
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 163 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g8m11.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 39-131

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
   1  aagggagctt atttgttcct tctatctatc atttatagAC ATAGCTGGAA
  51  GGCACCATCC ATGAACCAAC AAATAGGCTC TCACCAGAAA TCAAATATTC
 101  CTTGATCTTG GATTTTTCAG CCTCCAGAAC Tgtgagatat aaatttctgt
 151  tgtttacaag cta
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g9m21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
  1  TGATAGACTC ACAAGCCCTA TGTGGCATCA TGTAAGAATT ATCTTACTCT
 51  TGAACTGAG
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g9m21.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:61-119

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
  1  tatgcatgaa gaatcagtct ggataagagt gctcacttct gcctctactg
 51  tttctttcag TGATAGACTC ACAAGCCCTA TGTGGCATCATGTAAGAATT
101  ATCTTACTCT TGAACTGAGg tatggtgaag gaaatgccacatcattttgt
151  ttaatttacc aatatatctc caaataaact gcc
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g10m2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
  1  GGTCAGATGA AAGTGAGATC CATACATCCT TCTTCAGCAA CTTGTGCCTC
 51  TGCTCTGCAC CTCCCGCAAT TAACTACTGA AAAAAGAACA CAGCTTCACA
101  AAA
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g10m2.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 36-138

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
  1  gatcactatt ttcccctcca ctttaccccc tgcagGGTCA GATGAAAGTG
 51  AGATCCATAC ATCCTTCTTC AGCAACTTGT GCCTCTGCTC TGCACCTCCC
101  GCAATTAACT ACTGAAAAAA GAACACAGCT TCACAAAA gt gagttgaagt
151  gcatagcaca gaaccatcaa attccatatt tagaacaaga aaaattctat
201  aagacactat ttctg
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g11m3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
  1  GAGATTGTAA AATCAGGAAG TATATCTAAG TCACCTCCAG TAGCCGTAAC
 51  TCTACCTTGT CCAGTAAAAG
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 134 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g11m3.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 32-101

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
  1  tgtactatat gttttcatct tgtatttcta gGAGATTGTA AAATCAGGAA
 51  GTATATCTAA GTCACCTCCA GTAGCCGTAA CTCTACCTTG TCCAGTAAAA
101  Ggtaagtatg tgaaatagta taatttagaa gtaa
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 241 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g12m14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
   1  GGACAGGACC TCTATATGTT GCTCAGGCTG GAGTGTGGCA GCTATTCACG
  51  GATGTGATTA TAGTACACTA CAGCCTCAAC TCCTGGGCTC AAGCAAAGCT
 101  TCCCAAATCG GTGGGACTAT AGGCACACGC CACTGTGCTG TTCAATAATA
 151  AGATTTCTGT CTAACACCAC TGCGCCTGTT TCCTTGATAA ATATTTATTA
 201  TCTGTGTTTA TTTATTTTAT CCTGGGAGAA CATACAAGAT T
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 273 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g12m14.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 33-220

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
   1  ttctttttctt ttctttttctt tttaattttt agGGACAGGA CCTCTATATG
  51  TTGCTCAGGC TGGAGTGTGG CAGCTATTCA CGGATGTGAT TATAGTACAC
 101  TACAGCCTCA ACTCCTGGGC TCAAGCAAAG CTTCCCAAAT CGGTGGGACT
 151  ATAGGCACAC GCCACTGTGC TGTTCAATAA TAAGATTTCT GTCTAACACC
 201  ACTGCGCCTG TTTCCTTGAT aaatatttat tatctgtgtt tatttatttt
 251  atcctgggag aacatacaag att
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 432 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g13m22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
   1  GTTGTGAAAG AAAAATAAAT CTTGGGGCTC CAAAATCACT ACGCTAAAGG
  51  GAATAGTCAA GCTAGGAGCT GCTTACAGCA AACCTGCCTC CCATTCTATT
 101  CAAAGTCACC CCTCTGCTCA GAGATAAATG CATATCTGAT TGCCTCCTTT
 151  GGAGAGGCCA ATCAGAAACT CAAAAGAATG CAACTATTCA TCTCTTATCT
 201  ACCTATGACT TGGAAGCCCA CTCCCTGCTT CAAGTTGTCC CACCTTTGCT
 251  TCAAGTTGTC CAGCCTTTTC TGGACAGAAC CAGTGTTTAT CTTACATATA
 301  TTGACTGATG TCTCATGTCT CTCTAAAACA TATAAAACCA AGCTGTGCTC
 351  TTGAAGTGGC CTCATCATCT GGGGTGACAC CCGAGGTTCG TTGTCTCACG
 401  GCCATAAAGA TAAAGAACAT GGACACAAAA GG
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 702 bas pairs

(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g13m22.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:139-170

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
  1  atctaaagac atcactgggg gccatttgtc gacatgggtg ggccatttgt
 51  tgaagaaaag ttaaggtaag tatggttaaa tatttcttct tgaacttgtt
101  agattctcac agaaaatatt ctgttctctt gcctgcagGT TGTGAAAGAA
151  AAATAAATCT TGGGGCTCCA AAATCACTAC GCTAAAGGGA ATAGTCAAGC
201  TAGGAGCTGC TTACAGCAAA CCTGCCTCCC ATTCTATTCA AAGTCACCCC
251  TCTGCTCAGA GATAAATGCA TATCTGATTG CCTCCTTTGG AGAGGCCAAT
301  CAGAAACTCA AAAGAATGCA ACTATTCATC TCTTATCTAC CTATGACTTG
351  GAAGCCCACT CCCTGCTTCA AGTTGTCCCA CCTTTGCTTC AAGTTGTCCA
401  GCCTTTTCTG GACAGAACCA GTGTTTATCT TACATATATT GACTGATGTC
451  TCATGTCTCT CTAAAACATA TAAAACCAAG CTGTGCTCTT GAAGTGGCCT
501  CATCATCTGG GGTGACACCC GAGGTTCGTT GTCTCACGGC CATAAAGATA
551  AAGAACATGG ACACAAAAGG gtgaggttta gagcagaaat ttaataggtg
601  aaagaaagtg aacagctctc tgctacagag aggggtccca gaaaaatggg
651  ttgccgattc acagtttgga tacagaggct tttataagaa atcaatagtg
701  gg
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 128 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g14m12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
  1  CTATGGTAAC AGTTCCCTTC ATTACTTACT CAAACTTCAG AGAGATAAAA
 51  GAGAAGGAGT CACAGCATCT TTGTGCAAAA TATGCCTCGT TTTCTGGGAA
101  AAGGCTTGTT TCAGAAGAGA AGACAGTG
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g14m12.genomic

(ix) FEATURE:

(A) NAME/KEY:EXON: 37-164 87

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

1 cactgaatct gttctctgtg tttacttgta ttgtagCTAT GGTAACAGTT
51 CCCTTCAATA AAAGCACCAC CCTCAGAGAC AAAAAAGAGA AGGAGTCACA
101 GCATCTTTGT GCAAAATATG CCTCGTTTTC TGGGAAAAGG CTTGTTTCAG
151 AAGACACCGC tgctcttgtg gtgtcagaaa tgaggcaaat ggaacagagt
201 aatt

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 base pairs
(B) TYPE: nucleic acid
(ii) MOLECULE TYPE: cDNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(viii) NAME: c10r.c15m4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

1 GAGAATGCTG GCAGCTATTG AAAACAGACT GAAGAGAAGA AAGAAGATTT
51 GGAGGAATAT GTATCCAAGA TTGATGATGG TGGCTCAGAT CAG

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 165 base pairs
(B) TYPE: nucleic acid
(ii) MOLECULE TYPE: genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(viii) NAME: c15m4.genomic
(ix) FEATURE:
(A) NAME/KEY:EXON: 58-130
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

1 agcagaggaa ttacatgacc aactttcatt ttatcagGAT AATTCTGGCA
51 GTATTGAAA AAACTCTTAA GAAGAAAAC AAGATTTGGA CGAATATGTA
101 TCCAAGATTG ATGATGGTGG CTCAGATCAG gttggtagca gtgcagatgg
151 cgtttgggga tcaggt

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 121 base pairs
(B) TYPE: nucleic acid
(ii) MOLECULE TYPE: cDNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

Solutions of therapeutic compositions can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions also can be prepared in glycerol, liquid polyethylene glycols, mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The therapeutic compositions of the present invention are advantageously administered in the form of injectable compositions either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection also may be prepared. These preparations also may be emulsified. A typical composition for such purpose comprises a pharmaceutically acceptable carrier. For instance, the composition may contain 10 mg, 25 mg, 50 mg or up to about 100 mg of human serum albumin per milliliter of phosphate buffered saline. Other pharmaceutically acceptable carriers include aqueous solutions, non-toxic excipients, including salts, preservatives, buffers and the like.

Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oil and injectable organic esters such as ethyloleate. Aqueous carriers include water, alcoholic/aqueous solutions, saline solutions, parenteral vehicles such as sodium chloride, Ringer's dextrose, *etc.* Intravenous vehicles include fluid and nutrient replenishers. Preservatives include antimicrobial agents, anti-oxidants, chelating agents and inert gases. The pH and exact concentration of the various components the pharmaceutical composition are adjusted according to well known parameters.

An effective amount of therapeutic composition is determined based on the intended goal. The term "unit dose" or "dosage" refers to physically discrete units suitable for use in a subject, each unit containing a predetermined-quantity of the therapeutic composition calculated to produce the desired responses, discussed above, in association with its administration, i.e., the appropriate route and treatment regimen. The quantity to be administered, both according to number of treatments and unit dose, depends on the protection desired:

(viii) NAME: exon.g16m23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
  1  AACTCAGAAG TGATTCTCTA CTGAAATGAT ACAGTGTGCC CCAAAGACAT
 51  TGCATTAGAA CAATGGTAAT GGCAGGACTC TATGCCCATG ATTTCAGGAT
101  ACTTCTTCTA GGAATTAACT ATAT
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 324 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g16m23.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:89-212

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
  1  gtctatgtaa acacagaaaa aatataagaa ctaaccccag gcaatgactt
 51  ggtactgttt aaagaaagca ttcttcctta ttttctagAA CTCAGAAGTG
101  ATTCTCTACT GAAATGATAC AGTGTGCCCC AAAGACATTG CATTAGAACA
151  ATGGTAATGG CAGGACTCTA TGCCCATGAT TTCAGGATAC TTCTTCTAGG
201  AATTAACTAT ATgtaagtgt ctttttttatt gaaaatattg gactagctac
251  atcgagatgc ctttctgggt ttttttgcca ttagccaatt atgttagttt
301  tatgcttttg ctttatttca taag
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g17m24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
  1  ATTCTAGGGT ATCGATTCAT TTATAGTGGG GCAATCTTGC TGGAGATTAT
 51  AGAGAAAAG
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

WO 00/47773 PCT/US00/02052

85

treatment of certain disease states. One mechanism for delivery is *via* viral infection where the expression construct is encapsidated in an infectious viral particle.

4.8.5 Non-viral Methods

Several non-viral methods for the transfer of expression constructs into cultured mammalian cells also are contemplated by the present invention. These include calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe *et al.*, 1990); DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa *et al.*, 1986; Potter *et al.*, 1984), direct microinjection (Harland and Weintraub, 1985), DNA-loaded liposomes (Nicolau and Sene, 1982; Fraley *et al.*, 1979) and lipofectamine-DNA complexes, cell sonication (Fechheimer *et al.*, 1987), gene bombardment using high velocity microprojectiles (Yang *et al.*, 1990), and receptor-mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

In one embodiment of the invention, the expression construct may simply consist of naked recombinant vector. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. For example, Dubensky *et al.* (1984) successfully injected polyomavirus DNA in the form of $CaPO_4$ precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty and Neshif (1986) also demonstrated that direct intraperitoneal injection of $CaPO_4$ precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding an antisense UC41 construct also may be transferred in a similar manner *in vivo*.

Once the expression construct has been delivered into the cell the nucleic acid encoding the UC41 gene may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the gene may be stably integrated into the genome of the cell. This integration may be in the cognate location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the

---

(viii) NAME: g17m24.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:52-110

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
1    [illegible]at aaaatgtatc ttttttctc tctgaaaata
51   gATTCTAGGG TATCGATTCA TTTATAGTGG GGCAATCTTG CTGGAGATTA
101  TAGAGAAAAG gtcagttttc tgccaaatta ctttcattat gttgtccaag
151  aag[illegible]a
201  tatatctg
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g18m25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
1    AAACTTCTCA AAGACCCATC TACCCCTCAA TAGAAAGTAT CACCAGACAT
51   CAAAGCATCA AAG
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g18m25.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:94-156

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
1    tgatagcagg [illegible]
51   ttgagggacc ctcctgccca gacaaatgat ttgttcatta cagAAACTTC
101  TCAAAGACCC ATCTACCCCT CAATAGAAAG TATCACCAGA CATCAAAGCA
151  TCAAAGgtca gttttctgta aaattacttt tgttacattg tccctaaaaa
201  aaaaagacaa [illegible]
```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g19m5a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```
  1  TGCTGCTTGT TGCTGACGTT TCTGCCCAAA CATCATGAAA TGTGGAGAAG
 51  AAAGAACACT CAGATGTGCC CTGGAGACTT TCCTTTTATT TTTAAACAG
```

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g19m5b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
  1  TGCTGCTTGT TGCTGACGTT TCTGCCCAAA CATCATGAAA TGTGGAGAAG
 51  AAAGAACACT CAGATGTGCC CTGGAGACTT TCCTTTTATT TTTAAACAGG
101  TACATCCTGA AATTATGTTA GA
```

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 194 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g19m5.genomic

(ix) FEATURE:
(A) NAME/KEY:ALTERNATE EXON (a): 39-137
(A) NAME/KEY:ALTERNATE EXON (b): 39-160

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
  1  ttacataatc tcacttttтc ttattcgttt ccttttagTG CTGCTTGTTG
 51  CTGACGTTTC TGCCCAAACA TCATGAAATG TGGAGAAGAA AGAACACTCA
101  GATGTGCCCT GGAGACTTTC CTTTTATTTT TAAACAGGTA CATCCTGAAA
151  TTATGTTAGA gtaagtttca acatattcaa aatcatgtca taaa
```

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

components, is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are capable of infecting a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification could permit the specific infection of hepatocytes *via* sialoglycoprotein receptors.

A different approach to targeting of recombinant retroviruses has been designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, the infection of a variety of human cells that bear those surface antigens with an ecotropic virus *in vitro* was demonstrated (Roux *et al.*, 1989).

There are certain limitations to the use of retrovirus vectors. For example, retrovirus vectors usually integrate into random sites in the cell genome. This can lead to insertional mutagenesis through the interruption of host genes or through the insertion of viral regulatory sequences that can interfere with the function of flanking genes (Varmus *et al.*, 1981). Another concern with the use of defective retrovirus vectors is the potential appearance of wild-type replication-competent virus in the packaging cells. This may result from recombination events in which the intact

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon/genomic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
  1  GTTTTCTGCA [illegible] AATAGCATTT GAAGCTCATT GTCATTCGTG
 51  AATACATCTA GATTCACATG CTCAAAATCT AAAAGATTCA TAAGGATAAA
101  CCCTGGAAAG AACTCA[illegible] TCCTGGATAG CAGCCCTGAC CTACATGCTG
151  TTTAAATTAG TCTCTAGTCT TCAGCCCTCA CCAGTGCACA A
```

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 262 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: 130n5 genomic

(ix) FEATURE:
(A) NAME/KEY: EXON, 38-227

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```
  1  taaaatattt aactattttt gctttctttcc atctagATT TCCTCATCAC
 51  CTTTTTAATA TCATTTGAAT GTCATTCTCA TTCCTGAATA CATCTAGATT
101  CACATGCTCA AAATCTAAAA GATTCAGAAG GATAAACCCA GGAAAGACTC
151  TCTCCCTCTC GGATACCAGC CGTGACCTAC ATGCTGTTTA AATTAGTCTC
201  TAGTCTTCAG CCCTCACCAG TGCACAAgtg agtctggaac aactcaacc
251  cttattcctc ct
```

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon/genesis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```
  1  TTTTTTTTTC TTTTGGATTT GACGAAGTCT ATATTCGACT TGAACAAATC
 51  CGAAGGAAAG CTTCCAATTA TGGGGAACAA GTCCTCTGAA GTGGCTAAAT
101  TCCCAAAAAC AAAAAAAAAA TG
```

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 190 base pairs

(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g21m7.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON: 32-153

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```
  1  tggtttgttt ctgtttttgt tattgtttca gTTTTTTTTT CCATTGGGTT
 51  TGACCAACTC TATATTCGAC TTGAACAAAT CCGAAGGAAA GCTTCCAATT
101  ATGGGGAACA AGTCCTCTGA AGTGGCTAAA TTCCCACACA CACAAAAGAA
151  AAGgtagggt ggtgggggggg agaaaaacag ccagcaaaag
```

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 211 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g22m15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```
  1  ATGAGTCTAT GCCCAGGACC ACCAGATAAT TGAGTCCTGT ACAAAAGCTT
 51  CTGACTAAAC AATGTGCTCT GGCTCAGGAC TATACAGAGA AAAGACACAG
101  TTTTTAAATT GATCGTTCAA AAGGAAACAT ATTTATGATA TTTGCTCCAT
151  GATATGTATC TCTCATCTGT TAGCTCAGGC AGAATTAAAA TGCTAGACAA
201  TAAAAAAAAA A
```

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 375 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g22m15.genomic

(ix) FEATURE:
(A) NAME/KEY:EXON:52-254

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
  1  aatatgtaca tatcttagag aatcatttat tgtaattgtt ttcttttcca
 51  gATGAGTCTA TGCCCAGGAC CACCAGATAA TTGAGTCCTG TACAAAAGCT
101  TCTGACTAAA CAATGTGCTC TGGCTCAGGA CTATACAGAG AAAAGACACA
151  GTTTTTAAAT TGATCGTTCA AAAGGAAACA TATTTATGAT ATTTGCTCCA
201  TGATATGTAT CTCTCATCTG TTAGCTCAGG CAGAATTAAA ATGCTAGACA
```

```
  251  ATAAtgttcc atgccactat gcttactcgg gtctctcaca ttggtgtact
  301  tctggcacaa agtcacatga catttgagta atagctgtct cccatgtgga
  351  ctttacacca catgatgtca ggggc
```

medium is then replaced with 50 ml of fresh medium and shaking is initiated. For virus production, cells are allowed to grow to about 80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary overnight, following which the volume is increased to 100% and shaking is commenced for another 72 h.

Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain the conditional replication-defective adenovirus vector for use in the present invention. This is because Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

A typical vector applicable to practicing the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the UC41 gene at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical. The polynucleotide encoding the UC41 gene also may be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

Adenovirus is easy to grow and manipulate and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, e.g., 10^9-10^{11} plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g24m16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```
  1  AATTAAAACG ATGTATTAAG CTGGCCTTTT TTCAGACATA CTTGCAACCA
 51  GAGTCATCAA TTTTGAAGAC AGCAACCAAG CCAAGCAATG TGAGGTTACA
101  GCTATGAAAT AGAACAGAGA TGTCTAGACT ATAGACACAG CCTGCCGTTT
151  TGTGCTGATT GGTAAAGTGT TCCAGCCAAC TGGAAGCAAA TATTTCTCAG
201  AAGCAGTTTC CTGCTCTCAT CCTCTCCTCG CCATGCCCAC TGTGCCCAAC
251  ATGGCTCCAG CTGGGTCACA GAAGACTTTG TCCTGGAATA CAGCATTTCC
301  CTATTTAAAT CTCTAACTTT GTATGTACTC TTTTCAATAA AAGCATATTT
351  TTCATTACCA AAAAAAAAAA
```

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: g24m16.genomic

(ix) FEATURE:
(A) NAME/KEY: EXON:71-440

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

```
  1  gtatttatag tctgtccttc ttatgtagag aacacgaatg tctaattatt
 51  tacatttatt tttatgatag AATTAAAACG ATGTATTAAG CTGGCCTTTT
101  TTCAGACATA CTTGCAACCA GAGTCATCAA TTTTGAAGAC AGCAACCAAG
151  CCAAGCAATG TGAGGTTACA GCTATGAAAT AGAACAGAGA TGTCTAGACT
201  ATAGACACAG CCTGCCGTTT TGTGCTGATT GGTAAAGTGT TCCAGCCAAC
251  TGGAAGCAAA TATTTCTCAG AAGCAGTTTC CTGCTCTCAT CCTCTCCTCG
301  CCATGCCCAC TGTGCCCAAC ATGGCTCCAG CTGGGTCACA GAAGACTTTG
351  TCCTGGAATA CAGCATTTCC CTATTTAAAT CTCTAACTTT GTATGTACTC
401  TTTTCAATAA AAGCATATTT TTCATTACCA ttctgaccat actcccttct
```

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: exon.g25m8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

```
  1  TCTTCAAGAT AGAAAATAAG CTGTCTCTGA AGAGGTGCAA GTGGGATGCT
```

infection, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in humans.

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNAs for translation.

In currently used systems, recombinant adenovirus is generated from homologous recombination between shuttle vector and provirus vector. Due to the possible recombination between two proviral vectors, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

51 CCCAGGCATT TCTTGTAAGT TGTAATACCA CATTGCCTGA AATGATAGT

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 base pairs

(B) TYPE: nucl ic acid

(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(vii) [illegible]: 20mer [illegible]

(ix) FEATURE:

(A) NAME/KEY:EXON:36-174

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

1 tgagcaattt gttatgttca ttttttactt ttatagTCTTC AAGATAGAAA
51 ATAAGCTGTC TCTGAAGAGG TGCAAGTGGG ATGCTCCCAG GCATTTCTTG
101 TAAGTTGTAA TACCACATTG CCTGAAATGA TAGTGATATA TTCTTTATGT
151 AACTTTACAA ACTTGCAGGT TTTG

(2) INFORMATION FOR SEQ ID NO:53:

ACCGTAGCTACGTACGTATATAGAAAGGGCGCGATCGTCGTCGCGTATGACGACTTAGCATGC

(2) INFORMATION FOR SEQ ID NO:54:

ACCGGTAGCTACGTACGTTATTTAGAAAGGGGTGTGTGTGTGTGTGTAAACCGGGGTTTTCGGGATCGTCCGTCGCG
TATGACGACTTAGCCATGCACGGTATATCGTATTAGGACTAGCGATTGACTAG

(2) INFORMATION FOR SEQ ID NO:55:

ATGACCATAGTCGACCTGGCCGTCGTT

(2) INFORMATION FOR SEQ ID NO:56:

ATGAGCATAGTCGACGGATCCGTCGACCTGGCCGTCGTT

(2) INFORMATION FOR SEQ ID NO:57:

CGGAATTCTGCAGATCTA'B'C'NNNNNN

(2) INFORMATION FOR SEQ ID NO:58:

CGGAATTCTGCAGATCT[illegible]

(2) INFORMATION FOR SEQ ID NO:59:

CGGAATTCTGCAGATCT

(2) INFORMATION FOR SEQ ID NO:60:

(NH2)-GTAGTGCAAGGCTCGAGAACNNNNNNNNNNN

(2) INFORMATION FOR SEQ ID NO:61:

(NH2)-TGAGTAGAATTCTAACGGCCGTCATTGTTC

(2) INFORMATION FOR SEQ ID NO:62:

GAACAATGACGGCCGTTAGAATTCTACTCA-(NH2)

(2) INFORMATION FOR SEQ ID NO:63:

TGAGTAGAATTCTAACGGCCGTCAT

(2) INFORMATION FOR SEQ ID NO:64:

(PO4)-GTAGTGCAAGGCTCGAGAAC

(2) INFORMATION FOR SEQ ID NO:65:

(PO4)-TGAGTAGAATTCTAACGGCCGTCATTG

(2) INFORMATION FOR SEQ ID NO:66:

(NH_2)-GTAGTGCAAGGCTCGAGAACTTTTTTTTTTTTTT

(2) INFORMATION FOR SEQ ID NO:67:

(PO4)-GTAGTGCAAGGCTCGAGAACTTTT

(2) INFORMATION FOR SEQ ID NO:68:

(NH2)-GTAGTGCAAGGCTCGAGAACNNNNNNNNNNN

(2) INFORMATION FOR SEQ ID NO:69:

(NH2)-TGAGTAGAATTCTAACGGCCGTCATTGTTC

(2) INFORMATION FOR SEQ ID NO:70:

GAACAATGACGGCCGTTAGAATTCTACTCA-(NH2)

(2) INFORMATION FOR SEQ ID NO:71:

TGAGTAGAATTCTAACGGCCGTCAT

(2) INFORMATION FOR SEQ ID NO:72:

(PO4)-GTAGTGCAAGGCTCGAGAAC

(2) INFORMATION FOR SEQ ID NO:73:

(PO4)-TGAGTAGAATTCTAACGGCCGTCATTG 77

methods differ in their respective abilities to entrap aqueous material and their
(2) INFORMATION FOR SEQ ID NO:74:
respective aqueous space-to-lipid ratios.
AGG AAG TAT ATC TAA GTC ACC TCC A

The dried lipids or lyophilized liposomes prepared as described above may be
(2) INFORMATION FOR SEQ ID NO:75:
reconstituted in a solution of nucleic acid and diluted to an appropriate concentration
(Biotin)-AA TTC CAG ACA GAT TGC AGG CAC
with an suitable solvent, *e.g.*, DPBS. The mixture is then vigorously shaken in a vortex
(2) INFORMATION FOR SEQ ID NO:76:
mixer. Unencapsulated nucleic acid is removed by centrifugation at 29,000 × g and the
(PO4)-AG AGG ACT TGT TCC CCA TAA TTG
liposomal pellets washed. The washed liposomes are resuspended at an appropriate
(2) INFORMATION FOR SEQ ID NO:77:
total phospholipid concentration, *e.g.*, about 50-200 mM. The amount of nucleic acid
(Biotin)-AG AGG ACT TGT TCC CCA TAA TTG
encapsulated can be determined in accordance with standard methods. After
determination of the amount of nucleic acid encapsulated in the liposome preparation,
(2) INFORMATION FOR SEQ ID NO:78:
the liposomes may be diluted to appropriate concentration and stored at 4°C until use.
(PO4)-TT ACG GCT ACT GGA GGT GAC TTA

In a preferred embodiment, the lipid dioleoylphosphatidylcholine is employed.
(2) INFORMATION FOR SEQ ID NO:79:
Nuclease-resistant oligonucleotides were mixed with lipids in the presence of excess
(PO4)-AA GTC TCC AGG GCA CAT CTG A
t-butanol. The mixture was vortexed before being frozen in an acetone/dry ice bath.
(2) INFORMATION FOR SEQ ID NO:80:
The frozen mixture was lyophilized and hydrated with Hepes-buffered saline (1 mM
(Biotin)-AA AAA AAA AAA AAA AAA AA
Hepes, 150 mM NaCl, pH 7.5) overnight, and then the liposomes were sonicated in a
bath type sonicator for 10 to 15 min. The size of the liposomal-oligonucleotides
(2) INFORMATION FOR SEQ ID NO:81:
typically ranged between 200-300 nm in diameter as determined by the submicron
(PO4)-CGAAGGAAAGCTTCCAATTATG
particle sizer autodilute model 370 (Nicomp, Santa Barbara, CA).

(2) INFORMATION FOR SEQ ID NO:82:

4.8.4 Viral Delivery Systems

GTA ATG AAA TCT GAG AAG CTG AA

There are a number of ways in which expression vectors may introduced into
(2) INFORMATION FOR SEQ ID NO:83:
cells. In certain embodiments of the invention, the expression construct comprises a
CAC ACA GTG GTT AAT CAT AAA TAC
virus or engineered construct derived from a viral genome. The ability of certain
(2) INFORMATION FOR SEQ ID NO:84:
viruses to enter cells *via* receptor-mediated endocytosis, to integrate into a host cell
CAC AAA GGT ATC TTT TAA GTT CC
genome, and express viral genes stably and efficiently have made them attractive
(2) INFORMATION FOR SEQ ID NO:85:
candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988;
GTT TTC CCA GTC ACG ACG GAA GCT GAA TTT AGC AAT ACA GA
Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986; Temin, 1986). Preferred
gene therapy vectors are generally viral vectors.

(2) INFORMATION FOR SEQ ID NO:86:

GTT TTC CCA GTC ACG ACG TTA TCT GTT CAC TTC ACC TTT G

(2) INFORMATION FOR SEQ ID NO:87:

AGG AAA CAG CTA TGA CCA TCC TGA GCT TTC AAA AAA GTA TTC

(2) INFORMATION FOR SEQ ID NO:88:

AGG AAA CAG CTA TGA CCA TGG TCT TCA CTT TTC ATT TAC TTC

(2) INFORMATION FOR SEQ ID NO:89:

TAG CAT TGT TTG AAG CCA CAG

(2) INFORMATION FOR SEQ ID NO:90:

CTG GAA GAA ACC TGT AAC TTG

(2) INFORMATION FOR SEQ ID NO:91:

GTT TTC CCA GTC ACG ACG TGA AGC CAC AGA GTT TTA GAG

(2) INFORMATION FOR SEQ ID NO:92:

AGG AAA CAG CTA TGA CCA TTG TTC TCA AAT AAT GTC CCA AA

(2) INFORMATION FOR SEQ ID NO:93:

GTA ATG CTA TAA TGT TTG AAA GG

(2) INFORMATION FOR SEQ ID NO:94:

TTC AGG CTA ACT TCC ATC TTC

(2) INFORMATION FOR SEQ ID NO:95:

GTT TTC CCA GTC ACG ACG GGT TAC CCC AAC ATA CCT ATG

(2) INFORMATION FOR SEQ ID NO:96:

AGG AAA CAG CTA TGA CCA TAA ATA GCA TAC ATA ATG TTT ATT C

(2) INFORMATION FOR SEQ ID NO:97:

CAA AGA GTA TGG GAG GCT GA

(2) INFORMATION FOR SEQ ID NO:98:

ACT TCA GAG AAC AAC TTC GTC C

(2) INFORMATION FOR SEQ ID NO:99: 75

GTT TTC TCA GTC ACG ACG GGC TGA GAC TGA CTT GAC TAT T

(2) INFORMATION FOR SEQ ID NO:100:

AGG AAA CAG CTA TGA CCA TGA GGG TCC ATC AGC CTA C

(2) INFORMATION FOR SEQ ID NO:101:

GTG AAT GGC TAG ATC CCC TTT

(2) INFORMATION FOR SEQ ID NO:102:

AAT GAA CCT ACA GTG AGG CAG

(2) INFORMATION FOR SEQ ID NO:103:

GTT TTC CCA GTC ACG ACG AAC GAG TGT AAT GTG C

(2) INFORMATION FOR SEQ ID NO:104:

AGG AAA CAG CTA TGA CCA TGT TCT TTT ACA TCT TAA CCC AG

(2) INFORMATION FOR SEQ ID NO:105:

TCT AGT CAG CCT TCT TGA AC

(2) INFORMATION FOR SEQ ID NO:106:

GAC CTA AGA GCT CAA AGG AAT

(2) INFORMATION FOR SEQ ID NO:107:

GTT TTC CCA GTC ACG ACG CCT TCT TGA ACT AGA ACT TG

(2) INFORMATION FOR SEQ ID NO:108:

AGG AAA CAG CTA TGA CCA TCA GGG TTT ATC CTT ATG AA

(2) INFORMATION FOR SEQ ID NO:109:

GTT TTC CCA GTC ACG ACG TCA CAT GCT CAA AAT CTA AA

(2) INFORMATION FOR SEQ ID NO:110:

AGG AAA CAG CTA TGA CCA TAA CGG TAT CTC ACT TC

(2) INFORMATION FOR SEQ ID NO:111:

CTG AAT TGG CGT TTC GCT TC

nuclear non-histone chromosomal proteins (HMG-1) (Kato *et al.*, 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In that such expression vectors have been successfully employed in transfer and expression of a polynucleotide *in vitro* and *in vivo*, then they are applicable for the present invention. Where a bacterial promoter is employed in the DNA construct, it also will be desirable to include within the liposome an appropriate bacterial polymerase.

"Liposome" is a generic term encompassing a variety of single and multilamellar lipid vehicles formed by the generation of enclosed lipid bilayers. Phospholipids are used for preparing the liposomes according to the present invention and can carry a net positive charge, a net negative charge or are neutral. Dicetyl phosphate can be employed to confer a negative charge on the liposomes, and stearylamine can be used to confer a positive charge on the liposomes.

Lipids suitable for use according to the present invention can be obtained from commercial sources. For example, dimyristyl phosphatidylcholine ("DMPC") can be obtained from Sigma Chemical Co., dicetyl phosphate ("DCP") is obtained from K & K Laboratories (Plainview, NY); cholesterol ("Chol") is obtained from Calbiochem (La Jolla, CA); dimyristyl phosphatidylglycerol ("DMPG") and other lipids may be obtained from Avanti Polar Lipids, Inc. (Birmingham, Ala.). Stock solutions of lipids in chloroform, chloroform/methanol or *t*-butanol can be stored at about -20°C. Preferably, chloroform is used as the only solvent since it is more readily evaporated than methanol.

Phospholipids from natural sources, such as egg or soybean phosphatidylcholine, brain phosphatidic acid, brain or plant phosphatidylinositol, heart cardiolipin and plant or bacterial phosphatidylethanolamine are preferably not used as the primary phosphatide, *i.e.*, constituting 50% or more of the total phosphatide composition, because of the instability and leakiness of the resulting liposomes.

Liposomes used according to the present invention can be made by different methods. The size of the liposomes varies depending on the method of synthesis. A

(2) INFORMATION FOR SEQ ID NO:112:

AAA GAA AGC AGA ACC TTA GC

(2) INFORMATION FOR SEQ ID NO:113:

GTT TTC CCA GTC ACG ACG TTC TCC TTA CCA TTA GAG CA

(2) INFORMATION FOR SEQ ID NO:114:

AGG AAA CAG CTA TGA CCA TAT AGG TGG CCT TGT TAT GTA

(2) INFORMATION FOR SEQ ID NO:115:

TTC TCC TTA CCA TTA GAG CAC

(2) INFORMATION FOR SEQ ID NO:116:

[FAM]-CC TTC GGA TTT GTT CAA GTC

(2) INFORMATION FOR SEQ ID NO:117:

CCA TTT GCC TAA TGA ATG AA

(2) INFORMATION FOR SEQ ID NO:118:

GTC AGA AAA TCT TGG GTG TA

(2) INFORMATION FOR SEQ ID NO:119:

GTT TTC CCA GTC ACG ACG CTT AAG AAA GAG ATT GCC A

(2) INFORMATION FOR SEQ ID NO:120:

AGG AAA CAG CTA TGA CCA TGC AAT GTG GTA TTA CAA CTT A

(2) INFORMATION FOR SEQ ID NO:121:

GTT TTC CCA GTC ACG ACG AAA ATA AGC TGT CTC TGA AG

(2) INFORMATION FOR SEQ ID NO:122:

AGG AAA CAG CTA TGA CCA TGG GTG TAA AAT AAT TTC TGG

(2) INFORMATION FOR SEQ ID NO:123:

CGT CTT ACT CAG TTT TGT ATT CT

(2) INFORMATION FOR SEQ ID NO:124:

CAT CTA GAA GTA TGC ATT TGG TA

In certain embodiments of the invention, the delivery of a nucleic acid in a cell may be identified *in vitro* or *in vivo* by including a marker in the expression construct. The marker would result in an identifiable change to the transfected cell permitting identification of expression. Enzymes such as herpes simplex virus thymidine kinase (*tk*) (eukaryotic) or chloramphenicol acetyltransferase (CAT) (prokaryotic) may be employed.

One also may include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. For example, the SV40, β-globin or adenovirus polyadenylation signal may be employed. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

4.8.2 Single-chain Antibodies

In yet another embodiment, one gene may comprise a single-chain antibody. Methods for the production of single-chain antibodies are well known to those of skill in the art. The skilled artisan is referred to U.S. Patent 5,359,046, (incorporated herein by reference) for such methods. A single chain antibody is created by fusing together the variable domains of the heavy and light chains using a short peptide linker, thereby reconstituting an antigen binding site on a single molecule.

Single-chain antibody variable fragments (scFvs) in which the C-terminus of one variable domain is tethered to the N-terminus of the other *via* a 15 to 25 amino acid peptide or linker, have been developed without significantly disrupting antigen binding or specificity of the binding (Bedzyk *et al.*, 1990; Chaudhary *et al.*, 1990). These Fvs lack the constant regions (Fc) present in the heavy and light chains of the native antibody. Single-chain antibodies to the protein products of the UC41 gene are contemplated within the scope of the present invention.

(2) INFORMATION FOR SEQ ID NO:125:

GTT TTC CCA GTC ACG ACG TGA ATG TTA TTT TCT GGA AGG C

(2) INFORMATION FOR SEQ ID NO:126:

AGG AAA CAG CTA TGA CCA TTC AAA TAA GGT ATA AAG ACA GAG

(2) INFORMATION FOR SEQ ID NO:127:

GTT TTC CCA GTC ACG ACG AAT CCC TGA ATG GAT AGC ACC C

(2) INFORMATION FOR SEQ ID NO:128:

AGG AAA CAG CTA TGA CCA TAA ATC ACA AAA ATG TCT AAG GTT

(2) INFORMATION FOR SEQ ID NO:129:

CTG AAT CTC CCC TAT TAG AAG T

(2) INFORMATION FOR SEQ ID NO:130:

AAG GCC ATT AA GAG GTT CTT AG

(2) INFORMATION FOR SEQ ID NO:131:

GTT TTC CCA GTC ACG ACG GAG TTA CAT TCA TTT TTC GAG TC

(2) INFORMATION FOR SEQ ID NO:132:

AGG AAA CAG CTA TGA CCA TTT CAA GAC CAG CCT GAC CAA C

(2) INFORMATION FOR SEQ ID NO:133:

TCC [illegible] TCT CAA GCT

(2) INFORMATION FOR SEQ ID NO:134:

CAT AGA AAT TCT GAC CTA CCC A

(2) INFORMATION FOR SEQ ID NO:135:

GTT TTC CCA GTC ACG ACG CCA AGC TGA TGG TAT GTA GAG

(2) INFORMATION FOR SEQ ID NO:136:

AGG AAA CAG CTA TGA CCA TTG TAA ATG GAT CTT GAA GAT CAT

(2) INFORMATION FOR SEQ ID NO:137:

GCA CAG AGC ACA TTC TGG TGA

(2) INFORMATION FOR SEQ ID NO:138:

TCC CAA AGA AAA CTA CTA GCC

(2) INFORMATION FOR SEQ ID NO:139:

GTT TTC CCA GTC ACG ACG CTG ATG ATC ACA GTC TCT AAG

(2) INFORMATION FOR SEQ ID NO:140:

AGG AAA CAG CTA TGA CCA TCC AGC AAA GTT GTT GTT GGTT

(2) INFORMATION FOR SEQ ID NO:141:

AGA CAG TTG GTA TTT AGG GA

(2) INFORMATION FOR SEQ ID NO:142:

TCA TTA TTG CAT TTT CTG GA

(2) INFORMATION FOR SEQ ID NO:143:

GTT TTC CCA GTC ACG ACG AGC CAT TTT CCT CTC TCC A

(2) INFORMATION FOR SEQ ID NO:144:

AGG AAA CAG CTA TGA CCA TGG GCT TCT TTT CCA CTT CAA

(2) INFORMATION FOR SEQ ID NO:145:

CAA CCA AAC TAT TAT GAA ACC G

(2) INFORMATION FOR SEQ ID NO:146:

AGT GGG GAG CCA GTG CTG TTA

(2) INFORMATION FOR SEQ ID NO:147:

GTT TTC CCA GTC ACG ACG TTA TAA TAA TCA CTA GAG ATA GG

(2) INFORMATION FOR SEQ ID NO:148:

AGG AAA CAG CTA TGA CCA TAA TCT TGT ATG TTC TCC CAG G

(2) INFORMATION FOR SEQ ID NO:149:

TTG GTG GCA GTA GAC TGT GGT

(2) INFORMATION FOR SEQ ID NO:150:

GAC AGC TAT TAC TCA AAT GTC A

(2) INFORMATION FOR SEQ ID NO:151: 71

GTT TTC CCA GTC ACG ACG TAA GAT TTT GCT ACG CAA ACT GT

(2) INFORMATION FOR SEQ ID NO:152:

AGG AAA CAG CTA TGA CCA TAG AGA GCT CAG TAA GCA TAG T

(2) INFORMATION FOR SEQ ID NO:153:

TGG ACA AGT CAA TGC ACT ACT G

(2) INFORMATION FOR SEQ ID NO:154:

TGA TTT AAG CTG CCC AGA TTT C

(2) INFORMATION FOR SEQ ID NO:155:

GTT TTC CCA GTC ACG ACG TCT TCT TTA GTT GAG AGA ACC T

(2) INFORMATION FOR SEQ ID NO:156:

AGG AAA CAG CTA TGA CCA TGG AGA GAT GTT GAG CAC AGT

(2) INFORMATION FOR SEQ ID NO:157:

GTT TTC CCA GTC ACG ACG ACA GCT ATG AAA TAG AAC AGA G

(2) INFORMATION FOR SEQ ID NO:158:

AGG AAA CAG CTA TGA CCA TGC ATA CGT GCA GCA ACA GAG A

(2) INFORMATION FOR SEQ ID NO:159:

TTG GTC TCA GAA ATA ATC TTA CTG G

(2) INFORMATION FOR SEQ ID NO:160:

GGA TGT AGC ACC TTG AAA TCA TTC

(2) INFORMATION FOR SEQ ID NO:161:

GTT TTC CCA GTC ACG ACG AGC CTA TGG ATG TAT TTA TTC AGT TA

(2) INFORMATION FOR SEQ ID NO:162:

AGG AAA CAG CTA TGA CCA TGT TCC ATT CGT TTC CTA TCA TTA G

(2) INFORMATION FOR SEQ ID NO:163:

GGC AAA AAA ATC AAT AAT ATG

TABLE 2 (Continued)

	Mouse or Type I Collagen
	Glucose-Regulated Proteins (GRP94 and GRP78)
	Rat Growth Hormone
	Human Serum Amyloid A (SAA)
	Troponin I (TN I)
	Platelet-Derived Growth Factor
	Duchenne Muscular Dystrophy
	SV40
	Polyoma
	Retroviruses
	Papilloma Virus
	Hepatitis B Virus
	Human Immunodeficiency Virus
	Cytomegalovirus

(2) INFORMATION FOR SEQ ID NO:164:

CAT TGC CCA CCT GTC TAA C

(2) INFORMATION FOR SEQ ID NO:165:

GTT TTC CCA GTC ACG ACG AAG ATT GTT AAA TGC TAC TGC

(2) INFORMATION FOR SEQ ID NO:166:

AGG AAA CAG CTA TGA CCA TTA TCA CTA TTC CCC TTG GC

(2) INFORMATION FOR SEQ ID NO:167:

GGA ATG TGG AGT AAT GTA AAC

(2) INFORMATION FOR SEQ ID NO:168:

CAC CAT GTT GAA ATT AAG CAG

(2) INFORMATION FOR SEQ ID NO:169:

GTT TTC CCA GTC ACG ACG GTA ATT GTT GAT AGT CCT CTG

(2) INFORMATION FOR SEQ ID NO:170:

AGG AAA CAG CTA TGA CCA TCA TAA AAC CAA AGC ATC CG

(2) INFORMATION FOR SEQ ID NO:171:

ATT TGC TGT CAC ATT ACC CTG

(2) INFORMATION FOR SEQ ID NO:172:

CAG CCT GCC TGG GTG ACA G

(2) INFORMATION FOR SEQ ID NO:173:

GTT TTC CCA GTC ACG ACG TGT CAC ATT ACC CTG TTT ATC

(2) INFORMATION FOR SEQ ID NO:174:

AGG AAA CAG CTA TGA CCA TTA AGA AGA GGT GAT ATT ACT TAC

(2) INFORMATION FOR SEQ ID NO:175:

CTA TTG TAA TGA ATG CTG CTG

(2) INFORMATION FOR SEQ ID NO:176:

CAG AAG ATT ATC GTG GTC ATC

of a promoter that is regulated in response to specific physiologic signals can permit inducible expression of an inhibitory protein. For example, a nucleic acid under control of the human PAI-1 promoter results in expression inducible by tumor necrosis factor. Additionally any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) also could be used to drive expression of a nucleic acid according to the present invention. Use of a T3, T7 or SP6 cytoplasmic expression system is another possible embodiment. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

Tables 2 and 3 list several elements/promoters which may be employed, in the context of the present invention, to regulate the expression of the gene of interest. This list is not intended to be exhaustive of all the possible elements involved in the promotion of gene expression but, merely, to be exemplary thereof.

Enhancers are genetic elements that increase transcription from a promoter located at a distant position on the same molecule of DNA. Enhancers are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

Below is a list of viral promoters, cellular promoters/enhancers, and inducible promoters/enhancers that could be used in combination with the nucleic acid encoding a gene of interest in an expression construct (Table 2 and Table 3). Additionally, any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB)

(2) INFORMATION FOR SEQ ID NO:177:

GTT [illegible] AAG TGA GTC CTA ACC CTG

(2) INFORMATION FOR SEQ ID NO:178:

AGG AAA CAG CTA TGA CCA TCG TGG TCA TCA TAA ACT AAA TAC

(2) INFORMATION FOR SEQ ID NO:179:

AAC [illegible]

(2) INFORMATION FOR SEQ ID NO:180:

AGA TGA GCA GCC CAC TAT TG

(2) INFORMATION FOR SEQ ID NO:181:

GTT TTC CCA GTC ACG ACG CCA TTT GTT GAA GAA AAG TTA AG

(2) INFORMATION FOR SEQ ID NO:182:

AGG AAA CAG CTA TGA CCA TCA GAA AAG GCT GGA CAA CTT G

(2) INFORMATION FOR SEQ ID NO:183:

GTT TTC CCA GTC ACG ACG CAA CTA TTC ATC TCT TAT CTA CC

(2) INFORMATION FOR SEQ ID NO:184:

AGG AAA CAG CTA TGA CCA [illegible] TTG ATT TC

(2) INFORMATION FOR SEQ ID NO:185:

GAA [illegible]

(2) INFORMATION FOR SEQ ID NO:186:

TAT CTG AAA AAC TAA TAA GCC AG

(2) INFORMATION FOR SEQ ID NO:187:

GTT [illegible] CTT TCT ACT CAG AGT CTA TG

(2) INFORMATION FOR SEQ ID NO:188:

AG[illegible] TAG TAA CAT AAT TGG CTA ATG GC

(2) INFORMATION FOR SEQ ID NO:189:

CAG GAT TAT ACT TTC ACT CAA G

(2) INFORMATION FOR SEQ ID NO:190:

GAC ATT TAA CTT AAT TTC ACT TG

(2) INFORMATION FOR SEQ ID NO:191:

GTT TTC CCA GTC ACG ACG ATA GAC TCA AGA AAA ATG CTA AG

(2) INFORMATION FOR SEQ ID NO:192:

AGG AAA CAG CTA TGA CCA TCT CCT TGT TAT TTC TAA ACC AG

(2) INFORMATION FOR SEQ ID NO:193:

TTG TCT ACC TGA ACC CCG AG

(2) INFORMATION FOR SEQ ID NO:194:

CAA AAT GGG GCT TGA TTA GG

(2) INFORMATION FOR SEQ ID NO:195:

GTT TTC CCA GTC ACG ACG TAC CTT TCT GTG CGT GAT AGC

(2) INFORMATION FOR SEQ ID NO:196:

AGG AAA CAG CTA TGA CCA TTT AGG GCT CAA ACT GAA ATG G

(2) INFORMATION FOR SEQ ID NO:197:

AGCCATTTTCCTCTCTCCA

(2) INFORMATION FOR SEQ ID NO:198:

GTTTTCCCAGTCACGACGCCACCACATACCACACTTC

(2) INFORMATION FOR SEQ ID NO:199:

CAGAATCGCATCAGTAATAGA

(2) INFORMATION FOR SEQ ID NO:200:

GTTTTCCCAGTCACGACGTGAAGACCTCTTTGAATTATC

(2) INFORMATION FOR SEQ ID NO:201:

GAAGCTGTGTTCTTTTTTCA

(2) INFORMATION FOR SEQ ID NO:202:

AGGAAACAGCTATGACCATCTGTGTTCTTTTTTCAGTAGTTA

applied in much the same way as described for antisense polynucleotide. Ribozyme sequences also may be modified in much the same way as described for antisense polynucleotide. For example, one could incorporate non-Watson-Crick bases, or make mixed RNA/DNA oligonucleotides, or modify the phosphodiester backbone, or modify the 2'-hydroxy in the ribose sugar group of the RNA.

Alternatively, the antisense oligo- and polynucleotides according to the present invention may be provided as RNA via transcription from expression constructs that carry nucleic acids encoding the oligo- or polynucleotides. Throughout this application, the term "expression construct" is meant to include any type of genetic construct containing a nucleic acid encoding an antisense product in which part or all of the nucleic acid sequence is capable of being transcribed. Typical expression vectors include bacterial plasmids or phage, such as any of the pUC or Bluescript™ plasmid series or, as discussed further below, viral vectors adapted for use in eukaryotic cells.

In preferred embodiments, the nucleic acid encodes an antisense oligo- or polynucleotide under transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrase "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation.

The term promoter will be used here to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40 early transcription units. These studies, augmented by more recent work, have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of DNA, and containing one or more recognition sites for transcriptional activator or repressor proteins.

---

(2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 824 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: splice variant group 1 # 639.seq

(ix) FEATURE:
(A) NAME/KEY: [illegible]
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

```
  1  [illegible]
 51  GAAATATCTT ACCGAGCACT TTATTCAATC TAAATTTAAA TAGAAGTTTT
101  [illegible]
151  ACTGGAAGAA GGGTCAGATG AAAGTGAGAT CCATACATCC TTCTTCAGCA
201  [illegible]
251  ACAGCTTCAC AAAACTATGG TAACAGTTCC CTTCATTACT TACTCAAACT
301  [illegible]
351  TCGTTTTCTG GGAAAAGGCT TGTTTCAGAA GAGAAGACAG TGTGCTGCTT
401  GTTGCTGACG TTTCTGCCCA AACATCATGA AATGTGGAGA AGAAAGAACA
451  [illegible]
501  CCATTGGGTT TGACCAACTC TATATTCGAC TTGAACAAAT CCGAAGGAAA
551  [illegible]
601  CACAAAAGAA AAGATGAGTC TATGCCCAGG ACCACCAGAT AATTGAGTCC
651  [illegible]
701  AGAAAAGACA CAGTTTTTAA ATTGATCGTT CAAAAGGAAA CATATTTATG
751  [illegible]
801  AAATGCTAGA CAATAAAAAA AAAA
```

(2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 149 amino acids
(B) TYPE: peptide

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: splice variant group 1 # 639.pep

(ix) FEATURE:
(A) NAME/KEY:
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 3 20 28 38 43 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

```
1  MKVRSIHPSS ATCASALHLP QLTTEKRTQL HKTMVTVPFI TYSNFREIKE
```

```
  51  KESQHLCAKY ASFSGKRLVS EEKTVCCLLL TFLPKHHEMW RRKNTQMCPG
 101  DFPFIFKQFF FPLGLTNSIF DLNKSEGKLP IMGNKSSEVA KFPHTQKKR*
```

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 850 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: splice variant group 1 # 1407.seq

(ix) FEATURE:
(A) NAME/KEY: TRANSLATION START: 96
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 15 20 28 38 43 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

```
   1  GAAAAGTGAA GACCTCTTTG AATTATCTTA TTTCATTTGA CTATGTTCCT
  75  CCTGAGTCAC AAAAAAAGGA TGTTACAGCT ATTTTTCTTA AGCTGATGGG
 125  CCAAAAGATT GTAGGTCAGA TGAAAGTGAG ATCCATACAT CCTTCTTCAG
 175  CAACTTGTGC CTCTGCTCTG CACCTCCCGC AATTAACTAC TGAAAAAAGA
 225  ACACAGCTTC ACAAAACTAT GGTAACAGTT CCCTTCATTA CTTACTCAAA
 275  CTTCAGAGAG ATAAAAGAGA AGGAGTCACA GCATCTTTGT GCAAAATATG
 325  CCTCGTTTTC TGGGAAAAGG CTTGTTTCAG AAGAGAAGAC AGTGTGCTGC
 375  TTGTTGCTGA CGTTTCTGCC CAAACATCAT GAAATGTGGA GAAGAAAGAA
 425  CACTCAGATG TGCCCTGGAG ACTTTCCTTT TATTTTTAAA CAGTTTTTTT
 475  TTCCATTGGG TTTGACCAAC TCTATATTCG ACTTGAACAA ATCCGAAGGA
 525  AAGCTTCCAA TTATGGGGAA CAAGTCCTCT GAAGTGGCTA AATTCCCACA
 575  CACACAAAAG AAAAGATGAG TCTATGCCCA GGACCACCAG ATAATTGAGT
 625  CCTGTACAAA AGCTTCTGAC TAAACAATGT GCTCTGGCTC AGGACTATAC
 675  AGAGAAAAGA CACAGTTTTT AAATTGATCG TTCAAAAGGA AACATATTTA
 725  TGATATTTGC TCCATGATAT GTATCTCTCA TCTGTTAGCT CAGGCAGAAT
 775  TAAAATGCTA GACAATAAAA AAAAAA
```

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 157 amino acids
(B) TYPE: peptide

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: splice variant group 1 # 1407.pep

(ix) FEATURE:
(A) NAME/KEY:
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 15 20 28 38 43 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

affected. Similarly, detrimental non-specific inhibition of protein synthesis also can be measured by determining target cell viability *in vitro*.

As used herein, the terms "complementary" or "antisense" mean polynucleotides that are substantially complementary over their entire length and have very few base mismatches. For example, sequences of fifteen bases in length may be termed complementary when they have a complementary nucleotide at thirteen or fourteen positions out of fifteen. Naturally, sequences which are "completely complementary" will be sequences which are entirely complementary throughout their entire length and have no base mismatches.

Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology, but also contains a non-homologous region (*e.g.*, a ribozyme) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

As stated above, although the antisense sequences may be full length cDNA copies, or large fragments thereof, they also may be shorter fragments, or "oligonucleotides," defined herein as polynucleotides of 50 or less bases. Although shorter oligomers (8-20) are easier to make and increase *in vivo* accessibility, numerous other factors are involved in determining the specificity of base pairing. For example, both binding affinity and sequence specificity of an oligonucleotide to its complementary target increase with increasing length. It is contemplated that oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 or 100 base pairs will be used. While all or part of the gene sequence may be employed in the context of antisense construction, statistically, any sequence of 14 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence.

In certain embodiments, one may wish to employ antisense constructs which include other elements, for example, those which include C-5 propyne pyrimidines.

```
  1  MGQKIVGQMK VRSIHPSSAT CASALHIRQL TTEKRTQLHK TMVTVPFITY
 51  SNFREIKEKE SQHLCAKYAS FSGKRLVSEE KTVCCLLLTF LPKHHEMWRR
101  KNTQMCRGDF PFIFKQFFFP LGLTNSIFDL NKSEGKLPIM GNKSSEVAKF
151  PHTQKKR
```

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 980 base pairs
(B) TYPE: [illegible]
(ii) MOLECULE TYPE: [illegible]
(iii) HYPOTHETICAL: [illegible]
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(viii) NAME: splice variant group 2 # 1291.seq
(ix) FEATURE:
(A) NAME/KEY: TRANSLATION START: 168
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 3 20 28 38 43 47
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

```
  1  AAAAAAAGCC ATTTTCCTCT CTGGACCACA TACCACACTT CTGGGATTCT
 51  GAAATATCTT ACCGAGCACT TTATTCAATC TAAATTTAAA [illegible]
101  GCCCACTTCC CAAGAGAGAA ACAACAACGA ACTAGATGAG AATGAGAGGA
151  ACTGGAAGAA GGGTCAGATG AAAGTGAGAT CCATACATCC TTCTTCAGCA
201  ACTTGTGCCT CTGCTCTGCA CCTCCCGCAA TTAACTACTG AAAAAAGAAC
251  ACAGCTTGAG AAAACTATGG TAACAGTTCC CTTCATTACT TACTCAAACT
301  TCAGAGAGAT AAAAGAGAAG GAGTCACAGC [illegible] [illegible]
351  TCGTTTTCTG GGAAAAGGCT TGTTTCAGAA GAGAAGACAG TGTGCTGCTT
401  GTTGCTGACG TTTCTGCCCA AACATC[illegible] [illegible] [illegible]
451  CTCAGATGTG CCCTGGAGAC TTTCCTTTTA TTTTTAAACA GTTTTTTTTT
501  CCATTGGGTT TGACCAACTC [illegible] [illegible] [illegible]
551  GCTTCCAATT ATGGGGAACA AGTCCTCTGA AGTGGCTAAA TTCCCACACA
601  CACAAAAGAA AAGGTGCCTG [illegible] [illegible] [illegible]
651  TCTGGATTCA CTAAAACTAT ACTCTTAATC TGTACCGGGA AGTTCTATAC
701  TGTTGCTTTA ATCCTTGGTG [illegible] [illegible] [illegible]
751  GACAATCTGC CTGCAACACT CATAATCCCT GAATGGATAG CACCCAATAA
801  AAGAGAGAAC ATCATTGCCA [illegible] [illegible] [illegible]
851  CTTTATACCT TATTTGACCT GAATATCTTA TCTTTCCGTG TTAAATGCTC
901  TTCTCTCTTG ACCTCTAGAT [illegible] [illegible] [illegible]
951  CCCTAATATT TTCCGCTTTA AATTGTCCAC
```

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 193 amino acids
(B) TYPE: peptide
(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: YES
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(viii) NAME: splice variant group 2 # 1291.pep

(ix) FEATURE:
(A) NAME/KEY:
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 3 20 28 38 43 47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

```
  1  MKVRSIHPSS ATCASALHLP QLTTEKRTQL HKTMVTVPFI TYSNFREIKE
 51  KESQHLCAKY ASFSGKRLVS EEKTVCCLLL TFLPKHHEMW RRKNTQMCPG
101  DFPFIFKQFF FPLGLTNSIF DLNKSEGKLP IMGNKSSEVA KFPHTQKKRC
151  LQSVWNSDYY PSGFTKTILL ICTGKFYTVA LILGAVYFNQ DTP*
```

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1007 base pairs
(B) TYPE: nucleic acid

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(viii) NAME: splice variant group 2 # 2843.seq

(ix) FEATURE:
(A) NAME/KEY: TRANSLATION START: 96
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 15 20 28 38 43 47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

```
   1  GAAAAGTGAA GACCTCTTTG AATTATCTTA TTTCATTTGA CTATGTTCCT
 125  CCTGAGTCAC AAAAAAAGGA TGTTACAGCT ATTTTTCTTA AGCTGATGGG
 175  CCAAAAGATT GTAGGTCAGA TGAAAGTGAG ATCCATACAT CCTTCTTCAG
 225  CAACTTGTGC CTCTGCTCTG CACCTCCCGC AATTAACTAC TGAAAAAAGA
 275  ACACAGCTTC ACAAAACTAT GGTAACAGTT CCCTTCATTA CTTACTCAAA
 325  CTTCAGAGAG ATAAAAGAGA AGGAGTCACA GCATCTTTGT GCAAAATATG
 375  CCTCGTTTTC TGGGAAAAGG CTTGTTTCAG AAGAGAAGAC AGTGTGCTGC
 425  TTGTTGCTGA CGTTTCTGCC CAAACATCAT GAAATGTGGA GAAGAAAGAA
 475  CACTCAGATG TGCCCTGGAG ACTTTCCTTT TATTTTTAAA CAGTTTTTTT
 525  TTCCATTGGG TTTGACCAAC TCTATATTCG ACTTGAACAA ATCCGAAGGA
 575  AAGCTTCCAA TTATGGGGAA CAAGTCCTCT GAAGTGGCTA AATTCCCACA
 625  CACACAAAAG AAAAGGTGCC TGCAATCTGT CTGGAATTCA GACTATTACC
 675  CTTCTGGATT CACTAAAACT ATACTCTTAA TCTGTACCGG GAAGTTCTAT
 725  ACTGTTGCTT TAATCCTTGG TGCTGTTTAT TTCAACCAAG ATACACCTTG
 775  AAGACAATCT GCCTGCAACA CTCATAATCC CTGAATGGAT AGCACCCAAT
 825  AAAAGAGAGA ACATCATTGC CAATTTTTTC TTCTATCCAA GTCCTCCTCT
 875  GTCTTTATAC CTTATTTGAC CTGAATATCT TATCTTTCCG TGTTAAATGC
 925  TCTTCTCTCT TGACCTCTAG ATCACTGTAC TCTCTGGGGT CTGCCTCTGT
 975  TTCCCTAATA TTTTCCGCTT TAAATTGTCC ACA
```

(2) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 amino acids
(B) TYPE: peptide

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

63

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapi ns

(viii) NAME: splice variant group 2 # 2843.pep

(ix) FEATURE:
(A) NAME/KEY:
(B) LOCATION:
(C) EXON COMPOSITION (SEQ ID #S): 15 20 28 38 43 47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

1 MGQKIVGQMK VRSIHPSSAT CASALHLPQL TTERRTQLHK IMVTVFTITT
51 GNFREIKKKF SQHLGAKYAS FSGKRLVEEE KTVCCLLLTF LPKHHEMWRR
101 KNTQMCPGDF PFIFKQFFFP LGLTNSIFDL NKSEGKLPTM GNRSSEVARF
151 FHTQKKRQLQ SWWNSDYYPS GFTKTLLLIC TGKFYTWALI LGAVYFNQDT
201 P*

Inhibitors could potentially be designed for UC41. This is complicated by the fact that no specific function has been identified for this gene products, and no data is available on its three-dimensional structures.

Identification of protein function may be extrapolated, in some cases, from the primary sequence data, provided that sequence homology exists between the unknown protein and a protein of similar sequence and known function. Proteins tend to occur in large families of relatively similar sequence and function. For example, a number of the serine proteases, like trypsin and chymotrypsin, have extensive sequence homologies and relatively similar three-dimensional structures. Other general categories of homologous proteins include different classes of transcriptional factors, membrane receptor proteins, tyrosine kinases, GTP-binding proteins, *etc.* The putative amino acid sequences encoded by the prostate specific gene of the present invention may be cross-checked for sequence homologies *versus* the protein sequence database of the National Biomedical Research Fund. Homology searches are standard techniques for the skilled practitioner.

Even three-dimensional structure may be inferred from the primary sequence data of the encoded protein(s). Again, if homologies exist between the encoded amino acid sequences and other proteins of known structure, then a model for the structure of the encoded protein may be designed, based upon the structure of the known protein. An example of this type of approach was reported by Ribas de Pouplana and Fothergill-Gilmore (1994). These authors developed a detailed three-dimensional model for the structure of Drosophila alcohol dehydrogenase, based in part upon sequence homology with the known structure of 3-α, 20-ß-hydroxysteroid dehydrogenase. Once a three-dimensional model is available, inhibitors may be designed by standard computer modeling techniques. This area has been reviewed by Sun and Cohen (1993), herein incorporated by reference.

INTERNATIONAL SEARCH REPORT

Int. ational application No.
PCT/US99/19508

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/11, 15/00; C07K 5/00, 16/00; C07H 21/04

US CL : 536/23.1, 24.33; 530/300, 387.1; 435/440

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.33; 530/300, 387.1; 435/440

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, CANCERLIT, CAPLUS, EMBASE, MEDLINE, SCISEARCH
search terms: hpc1, prostate, dna, mutation, protein, antibody

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	INOUE et al. Neuron specific expression of a membrane protein, HPC-1: tissue distribution, and cellular and subcellular localization of immunoreactivity and mRNA. Molecular Brain Research. 1993, Vol. 19, pages 121-128.	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92
A	EELES et al. Linkage analysis of chromosome 1q markers in 136 prostate cancer families. Am. J. Hum. Genet. March 1998, Vol. 62, No. 3, pages 653-658.	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92
A	DE LA CHAPELLE et al. The genetics of hereditary common cancers. Current Opinion in Genetics and Development. June 1998, Vol. 8, No. 3, pages 298-303	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 07 DECEMBER 1999	Date of mailing of the international search report 07 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN SIU Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

61

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	NOTO et al. A novel human RasGAP-like gene that maps within the prostate cancer susceptibility locus at chromosome 1q25. FEBS Letters, 11 December 1998, Vol. 441, pages 127-131.	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92
A, P	DENSMORE et al. Allelic imbalance in familial and sporadic prostate cancer at the putative human prostate cancer susceptibility locus, *HPC1*. British Journal of Cancer. December 1998, Vol. 78, No. 11, pages 1430-1433.	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92
A	GOONEY et al. Prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. Journal of the National Cancer Institute. 02 July 1997, Vol. 89, No. 13, pages 955-959.	8, 12-13, 19-24, 26, 28-44, 46-69, and 71-92

cells, but occurs at reduced levels in transformed cells, the "background" signal from the normal cells in the sample may mask the presence of transformed cells.

The ideal cancer marker would be one that is present in malignant cancers, and either missing or else expressed at significantly lower levels in benign tumors and normal cells. The present invention addresses this need for prostate cancer markers by identifying a novel, prostate specific gene (UC Band #41) which is expressed at much higher levels in malignant prostate carcinoma than in benign or normal prostate. In particular, the results for UC Band #41 (SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:4) discussed in Examples 1-5 below, are quite promising in that this marker is apparently only overexpressed in malignant tumors and is present at very low levels in BPH or normal prostate. Further, this gene is significantly elevated in a high percentage of human prostate cancers examined to date.

It is expected that in clinical applications, human tissue samples will be screened for the presence of the expression products of UC41. Such samples could consist of needle biopsy cores, surgical resection samples, lymph node tissue, or serum. In certain embodiments, nucleic acids would be extracted from these samples and amplified as described above. Some embodiments would utilize kits containing pre-selected primer pairs or hybridization probes. The amplified nucleic acids would be tested for UC41 expression products by, for example, gel electrophoresis and ethidium bromide staining, or Southern blotting, or a solid-phase detection means as described above. These methods are well known within the art. The levels of expression product(s) detected would be compared with statistically valid groups of metastatic, non-metastatic malignant, benign or normal prostate samples. The diagnosis and prognosis of the individual patient would be determined by comparison with such groups.

Another embodiment of the present invention involves application of RT-PCR™ techniques to detect circulating prostate cancer cells (*i.e.*, those that have already metastasized), using probes and primers selected from sequences or their complements designated herein as SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:4. Similar techniques

relative quantitative RT-PCR™ with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies described below were performed using a more conventional relative quantitative RT-PCR™ with an external standard protocol. These assays sample the PCR™ products in the linear portion of their amplification curves. The number of PCR™ cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This is very important since this assay measures absolute mRNA abundance. Absolute mRNA abundance may be used as a measure of differential gene expression only in normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCR™ assays may be superior to those derived from the relative quantitative RT-PCR™ with an internal standard.

One reason for this is that without the internal standard/competitor, all of the reagents may be converted into a single PCR™ product in the linear range of the amplification curve, increasing the sensitivity of the assay. Another reason is that with only one PCR™ product, display of the product on an electrophoretic gel or some other display method becomes less complex, has less background and is easier to interpret.

4.7 Diagnosis and Prognosis of Human Cancer

In certain embodiments, the present invention allows the diagnosis and prognosis of human prostate cancer by screening for prostate specific nucleic acids, particularly those that are overexpressed in prostate cancer. The field of cancer diagnosis and prognosis is still uncertain. Various markers have been proposed to be

INTERNATIONAL SEARCH REPORT

Internatio[n] application No.
PCT/US99/19508

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-7, 9-11, 14-18, 25, 27, 45, and 70
because they relate to parts of the international application that do not compl[y] with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Claims are drawn to sequence IDs, however, no computer readable format was submitted.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.